Three Essays on International Trade and Development Economics

by

Luis Alberto Baldomero Quintana

A dissertation presented in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Economics)
in the University of Michigan
2020

Doctoral Committee:

Professor C. Hoyt Bleakley Jr., Co-Chair
Professor Andrei Levchenko, Co-Chair
Professor Jagadeesh Sivadasan
Assistant Professor Sebastian Sotelo

Luis Alberto Baldomero Quintana

luisbq@umich.edu

ORCID iD: 0000-0002-3133-4754

A mi madre y a mi abuela, por enseñarme tenacidad y fortaleza.

Acknowledgements

The past seven years have been an authentic adventure. When I started this Ph.D. program back in 2013, I did not imagine the set of experiences, challenges and joyful moments I would live. I was blessed with a large group of people who supported me. “It took a village” to complete my graduate studies.

My committee members did more than only teaching me about economics. My co-chair Hoyt Bleakley has influenced enormously the way I think about economic development. He made sure I understood the context of a country extremely well (a unique feature among development economists), and he pushed me towards thinking about novel and different ideas in rigorous ways. He also supported me during one of the toughest moments of the program and he was very patient with my failures (probably more patient than myself). He taught me patience and to behave with courtesy all the time. My co-chair Andrei Levchenko was an impressive advisor. He helped me to clarify my ideas and to think about the main message of a research paper. He taught me to be organized, to be disciplined, and to communicate well. On a personal level, he taught me to listen to others, and to be honest about what I knew and what I didn’t. Sebastian Sotelo has been an incredible source of support since day one in the good and bad days, and I am completely sure I would have left the Ph.D. if I would not have found him as an advisor early in the program. He patiently walked me through every technical detail of international trade theory and listened to every single one of my research ideas, always providing detailed feedback. More importantly, he taught me the importance of being kind and professional, always leading by example. Jagadeesh Sivadasan taught me to be practical and rigorous at the same time. Visiting his office always implied learning new things about economics, and receiving extensive professional advice, something I really enjoyed. He taught me to evaluate with detail every idea. Even when he provided tough feedback, he was always very friendly. He constantly pushed me towards the final goal, and this is something that helped me to finish my job market paper on time.

I am also indebted to other people at Michigan Economics, Adam Stevenson taught me how to teach economics. He was a fantastic boss and the door of his office was always open to me. Working with him for five years was the best teaching job any graduate student can have. Also, I will be always indebted to Linda Tesar. She made sure I stayed in the Ph.D. program, and I will always be thankful to her for that. Her phrase since first year "We want you to succeed" kept me going during difficult times. Also, I do not have enough words to thank Laura Flak and Lauren Pulay for all their support. Their kindness is exceptional.

My close friends were essential for me to finish this adventure. Luis Alejos treated me like I was part of his family. Benjamin Glass was always there for me. From distance, four friends were key during my Ph.D. studies: visiting Joshua Francia and Adalberto Bustamante at the end of the academic year was a gift, and the constant communication with Abraham Aldama and Luis Monroy-Gomez-Franco was a constant source of energy and inspiration. The research group with Meera Mahadevan, Pinghui Wu and Oxana Azgaldova was key for me during multiple parts of the dissertation process. Vybhavi Balasundharam provided me with tons of support during, and after the job market. The advice of Jorge Luis García was key for me to finish my job market paper.

I received spiritual support from the catholic community in St. Mary's Student Parish. Specially Father Joe Wagner and Eric Sundrup. Every weekend, the community helped me. I spent a lot of Sundays in the study lounge as a way to recharge my spirit.

I am indebted to two Colombians who taught me every detail about their country: Maria Aristizabal and Wilmer Henao. Many of my good memories include multiple coffees with them. Also, hanging out with them was always a lot of fun. My best memories from the Ph.D. include multiple gatherings with Lorch 109 and company": Luis and Christina, Giacomo and Giulia, Owen, Will and Claire, Ben and Meghan. Moreover, I will always remember with joy all my Latin American friends in the Economics department, their presence helped me to enjoy my life as a graduate student at Michigan. I am also thankful that early in the program, I interacted with two economists who influenced me more than they know: Vanessa Alviarez and Monica Hernandez represent the economist I aspired to be. Seven Mexican economists supported me professionally and academically: Alexis Milo, Marco Oviedo, Lorely Ochoa, Jose de Jesús Salazar, Amado Villarreal, and Alejandro Ybarra-Yúnez, and Héctor Villarreal.

Ann Arbor was my home because of all the friends I made. I have excellent memories of the town because of them: Alecia Cassidy, Wenting Song, Xing Guo, Hannah Bolder, Guangye Cao, Pamela Martínez, Longxui Tian, Andrew Usher, Dyanne Vaught, Stephanie Karol, Ellie Epskamp-Hunt, Brian Cevallos, Sara del Fierro, Nadim Elayan Balagué, Alberto Arredondo, Rosina Rodríguez, Luis Espinoza, Yenibel Cuevas, Paula Ávila, Jenny Mayo, Thomas Helgerman, Ojaswi Adhikari, Travis Triggs, Claudia Villoro, Luis Guevara, Nabil López, Pamela Martínez, Barthelemy Bonadio, Tereza Ranozova, Magali Campos, Mariana Masteling, and Jack.

My friends from Mexico always believed in me and their support was very important: Alejandro Marín, Paul Feldman, Ismael Mireles, Adolfo Carranza, Erika González, Gerardo Márquez, Eduardo Smith, Efren Domínguez, Xavier Campos, Adrián Cuadros, Carlota Hervel, Francisco Parra, José Hernández, Isabel Sandoval.

Finally, I am thankful for the support of my mother Gloria, my brothers Juan and Edmundo, my grandmother Juana, my aunts Rosa, Josefina and Norma, and my cousin Óscar. Their love kept me going. Thank you for believing in this project.

Preface

This dissertation consists of three independent essays related to international trade in developing countries. I focus on the context of Colombia in all the essays. Firstly, I analyze how infrastructure is a determinant of comparative advantage. Secondly, I study the economic geography of commodity booms. Lastly, I examine the dynamics of importers after a trade liberalization event.

The first essay examines the effects of infrastructure projects on the composition of the basket of exports. I develop a model of domestic and international trade with different shipping routes and input-output linkages, to study how a large road infrastructure project affects the composition of the basket of exports of Colombia, a country with a historical concentration on commodity exports. I estimate parameters and calibrate the model using customs administrative data, a transportation survey and a unique dataset of geospatial data generated from physical road maps. Results indicate that improving the connectivity of a large manufacturing region to ports reduces Colombia's export concentration on commodity products, thus altering specialization. This suggests that infrastructure development policy could be an effective alternative to existing piecemeal government programs such as subsidies to specific sectors that aim to change comparative advantage.

The second essay studies the short-run spatial economic effects of commodity booms. Using the oil boom across Colombian regions as an example, I provide empirical evidence that input-output linkages, the composition of the local industry and domestic trade costs strongly influence the size and direction of the regional economic impacts of commodity booms. In addition, the transportation costs across different manufacturing goods matter when we measure the impact of commodity booms in the local industries, that is, sectors that produce goods that are difficult to transport due to physical characteristics experience a positive impact during a commodity boom. My results indicate that to generate a more precise picture of the impacts of commodity booms, it is necessary to consider factors related to general equilibrium effects. Specifically, industry linkages, domestic trade costs and the composition of the local industry have the capacity to change the regional economic impacts of commodity booms.

The third essay (joint with Vybhavi Balasundharam) studies the dynamics of importers, and whether these dynamics change after a trade liberalization event. Using detailed customs administrative data we provide evidence of three novel patterns among firms that import: the existence of churning of importers, convergence of new importers with respect to existing ones, and divergence across existing importers. We evaluate whether an FTA is responsible of these patterns and we find that this is not the case. The patterns seem to have specific trends across time, and the FTA does not seem to impact such trends. The results have implications to understand how changes in the composition of imports occur, and our results can help to understand better how imports impact productivity gains.

Table of Contents

List of Figures

List of Tables

List of Appendices

Abstract

My dissertation studies three topics related to international trade in a developing country context: infrastructure and trade, the geographical impact of commodity booms, and the dynamic behavior of importers. These topics are key to understanding how globalization impacts the development path of developing nations. For all my chapters, I use Colombia as a context. Firstly, I analyze how roads impact the type of goods a country exports. I focus on infrastructure given that middle and low income nations typically have low quality of road systems. Secondly, I focus on how commodity booms, economic events that are common in developing nations, shape the local manufacturing across regions in a country. Lastly, I analyze trade liberalizations from a new perspective: the dynamics of the importers. Understanding importers is key to analyze the gains of trade in developing nations because often developing nations take advantage of free trade agreements by importing high quality of inputs and capital goods.

The first chapter provides evidence that the road system of a country represents a source of comparative advantage. To do this I developed a model of international trade and internal geography, where regions within a country trade with each other and the rest of the world. The model include two relevant features: input-output linkages and multiple shipping routes. To quantify the model I use Colombian data including customs administrative data, a transportation survey, a geospatial data regarding the Colombian highway system based on physical road maps. The results of my calibration indicate that a large infrastructure project that connects the capital of the country with the Atlantic seaports would lead to a higher level of manufacturing exports. The results are highly relevant for a country like Colombia, a country that experiences a high concentration of

exports in commodities. I provide evidence that roads can help to diversify the export basket of a country.

The second chapter of my dissertation focuses on the economic geography of commodity booms. A large literature in economics has analyzed the regional impacts of commodity booms, but previous empirical work has not paid attention to general equilibrium effects. Using the oil boom across Colombian departments as an example, I provide evidence that the local impacts of commodity booms strongly depend on their size and sign on three general equilibrium elements: the composition of the local industry, the level of domestic trade costs and the input-output linkages. Moreover, these results also indicate that the complexity of the transportation of goods matters when we measure the local impacts of commodity booms in manufacturing outcomes. My empirical results suggests that the negative regional effects of natural resource booms documented in the resource curse literature could be alleviated and might change in the presence of industry linkages, high domestic trade costs, and specific configurations of the local industry.

The third chapter of my dissertation was written jointly with Vybhavi Balasundharam. We document different patterns of the dynamic of importers and we evaluate whether these patterns are impacted by a trade liberalization event. The patterns observed in the data are the following: the existence of churning among importers, a fast convergence of importers with respect to old importers across different years, and the divergence among existing importers on the intensive and extensive margins of trade. Such patterns do not seem to change after a the implementation of a Free Trade Agreement between Colombia and the United States. Our results help to understand the productivity gains generated by imports and how changes in the composition of imports are driven by specific firms.

Chapter 1. How Infrastructure Shapes Comparative Advantage

1.1. Introduction

Comparative advantage is a fundamental idea in international trade theory. Standard trade models typically examine the role of technology, institutions, and factor endowments to explain the patterns of specialization. However, this approach is limited by the fact that we only observe the patterns of international trade generated by regions within countries well connected to the global markets. This is especially true for developing nations, as the quality of infrastructure varies within these countries (Oxford Economics, 2017; IADB, 2013). Whether domestic trade costs within a country influence comparative advantage has not been studied in the literature.

This paper shows that domestic trade costs are indeed determinants of comparative advantage in a developing country context. As new infrastructure projects change the structure of the national transportation network or how industry linkages propagate shocks across regions and sectors, it is necessary to use a quantitative model to understand the mechanisms by which changes in domestic trade costs affect comparative advantage. Therefore, I build an international trade and internal geography model with input-output linkages, road networks, and international shipping routes. I use the model to understand the effects of completing a large infrastructure project currently in construction (*Ruta del Sol*) on the comparative advantage of Colombia. I show that the completion of the project increases the share of manufacturing exports and reduces the share of mining exports. Therefore, the highway project shifts the comparative advantage of Colombia away from the mining sector and towards manufacturing products.[1]

Colombia is an ideal context to analyze the impact of infrastructure on comparative advantage because the country is similar to several developing nations along a number of dimensions. First, Colombia's exports are concentrated in a few goods, particularly mining products. Second, there is variation in the access to global markets among Colombian departments.[2] In many developing countries there is a similar situation, with some regions with excellent access to global markets and others that are almost isolated due to poor infrastructure. Third, there is heterogeneity in the comparative advantage of Colombian regions. Many large middle-income nations share this characteristic.

I develop a framework in which departments in Colombia trade with each other and with the rest of the world. The model includes input-output linkages between three tradable sectors (agriculture, mining, and manufacturing) and a non-tradable sector (services). This characteristic allows trade costs to affect both output prices and production costs. Lastly, I include a realistic transportation feature: the existence of different shipping routes when departments and the rest of the world trade with each other. The model produces a tractable expression for the international trade flows between a department and the rest of the world, that use specific ports of exit or entry (a department-port gravity equation).

To take the model to the data, I combine four data sources: detailed customs administrative data with information about the port of exit or entry, a survey of transportation flows, and geospatial data that I create using digital and scanned physical road maps. The customs data allow me to obtain international trade flows between departments and the rest of the world, with information about the port used for exit or entry. The transportation survey allow me to obtain a proxy of domestic sectoral trade flows. Finally,

[1]Throughout the paper, I measure the comparative advantage of Colombia in a sector by using the share of exports. This works as a proxy to measure comparative advantage because, when the Balassa Index of Revealed Comparative Advantage is used for small open economies and highly aggregated sectors, the denominator of the index is fixed. French (2017) documents that revealed comparative advantage is useful to analyze the patterns of comparative advantage for different economies.

[2]A department is the official administrative region of Colombia, similar to states in the United States of America.

using the geospatial data and the Dijkstra's algorithm, I obtain travel times between any location within Colombia for both modern and historical road networks.

There are two parameters that govern my model. The first parameter defines the relationship between trade costs and travel times, and the second parameter defines the heterogeneity of the use of shipping routes for goods traded between Colombian regions and the rest of the world. To recover the values of these parameters, I estimate a department-port gravity equation using an instrumental variable approach. My instrument is the distance between locations using historical road networks during periods in which the characteristics of the Colombian economy were very different compared to the current economic circumstances (this approach is similar to Baum-Snow, 2007; Michaels, 2008; Duranton, Morrow, and Turner 2014; and Duranton, 2015). After obtaining the value of the parameters of my model, I run counterfactual simulations.

My main counterfactual experiment considers the effects of the infrastructure program *Ruta del Sol* on the sectoral exports of Colombia. The project's objective is to modernize the highway that connects Bogota with the Atlantic seaports. This department is the main exporter of manufacturing and agricultural products. Given the structure of the road system in Colombia, *Ruta del Sol* also improves substantially the access to international markets for several departments that specialize in the mining sector. Hence, the expected effect of this highway project in the national sectoral exports is unclear a priori. Additionally, given the structure of the input-output linkages in Colombia, the benefits of the reduction in domestic trade costs propagate in such a way that one sector benefits more than others.

The results of my counterfactual experiment show that the completion of the infrastructure project increases the share of manufacturing exports by four percentage points. To understand the importance of this result, note that for the past three decades, the share of mining exports of Colombia has grown substantially. This result implies that the road project can potentially reverse the upward trend of the specialization of Colombia in mining goods and shift the comparative advantage of Colombia towards the manufacturing sector. This result does not imply that the non-manufacturing exports fall, but rather the manufacturing exports grow more than the exports of other sectors. So, the infrastructure project can potentially revert any existing crowd-out effects of the mining boom on the Colombian manufacturing sector, due to potential Dutch disease effects (Alcott and Keniston, 2018).

To analyze the main forces driving my results, I run alternative counterfactual exercises, in which I isolate the different effects of *Ruta del Sol*. I consider separately the effects of the road project on domestic trade costs, international trade costs, and on both domestic and international trade costs without including input-output linkages. My alternative simulations show that industry linkages help to increase the manufacturing exports substantially. When I simulate the effects of *Ruta del Sol* without industry linkages, the increase in the share of manufacturing exports is one third of the growth observed in my main counterfactual experiment, which does consider these linkages. This is due to the fact that the manufacturing sector benefits more from access to tradable intermediate inputs, compared to the mining sector.

My work contributes to the international trade literature on the determinants of comparative advantage. My main contribution is to show that domestic trade costs are a source of national comparative advantage. This finding is specially relevant in developing countries where domestic trade costs are high, thereby generating differences in regional access to global markets within a country (Atkin and Donaldson, 2015). To my knowledge, recent international trade literature has provided little attention to the direct link that exists between the spatial distribution of domestic trade costs and national comparative advantage.

The main idea of this paper, how internal trade costs shape comparative advantage, is related to Deardoff (2014). He shows that the transportation costs of a nation to those countries that are geographically close impacts its comparative advantage. He defines the term *local comparative advantage*, which measures comparative advantage considering such transportation costs. With this term, it is possible to explain situations in which a country has a comparative advantage in a specific sector, even though production costs are high. In such cases, the comparative advantage exists due to low transportation costs between the economy and its neighboring nations. I focus exclusively on how the comparative advantage of a country is shaped exclusively by its internal transportation costs, while Deardoff (2014) focuses on transportation costs to the neighboring economies.

The closest works to this paper, are Duranton, Morrow, and Turner (2014) and Duranton (2015). These papers use applied microeconomics methods to show that urban centers with better infrastructure can specialize in sectors that produce heavy goods. Unlike these papers, I focus on how roads affect specialization at a national level. Additionally, I differ from such work by using an international trade model to run counterfactual scenarios that examine how a large infrastructure project can change comparative advantage. Besides, my theoretical framework considers the role of industry linkages.

Other work related to the determinants of comparative advantage includes papers regarding how migration affects specialization (Arkolakis, Lee and Peters, 2018; Bahar and Rapoport, 2018; Morales, 2019; Pellegrina and Sotelo, 2019), how the quality of institutions is a source of comparative advantage (Levchenko, 2007) or how domestic trade costs influence crop choices in developing countries (Allen and Atkin, 2018; Morando, 2019). This paper also speaks to the theoretical research regarding the dynamics of comparative advantage (Matsuyama, 1992; Krugman, 1987; Levchenko and Zhang, 2016; Hanson, Lind, and Muendler 2015).

In the international trade literature, there is an increasing interest in the effects of infrastructure projects. This includes work on how infrastructure improvements affect either domestic outcomes, or trade flows between a country and the rest of the world (Alder, 2019; Allen and Arkolakis 2019; Coatsworth 1979; Cosar and Demir, 2016; Ducruet et al, 2019; Donaldson 2018; Donaldson and Hornbeck, 2016; Faber, 2014; Fogel 1962; Holl, 2016; Perez-Cervantes, 2014; Xu, 2016; Xu 2018). To my knowledge, only two papers consider jointly domestic outcomes and international trade: Fajgelbaum and Redding (2018) on the structural transformation of Argentina during the period 1869-1914, and Sotelo (2019) on how roads affect agricultural trade in Peru. I depart from the existing literature by highlighting the role of industry linkages when I examine the effects of infrastructure in economic outcomes. Specifically, I show that input-output linkages propagate the effects of lower domestic trade costs. Although the previous work on infrastructure considers the effects of domestic trade costs on exports by sector, the interactions between industry linkages and infrastructure have not been examined in detail. However, to understand the impact of infrastructure on sectoral exports, it is crucial to consider the industry linkages. This is because the existence of such linkages generates uneven effects of changes in domestic trade costs on exports across sectors.

Lastly, my results are relevant for the literature on the Dutch disease. The work on this topic has been extensive, as Van der Ploeg (2011) points out. My results show that improving the domestic integration of regional markets in specific ways can minimize the specialization of an economy in a single sector. More precisely, my paper is closely related to one of the main mechanisms of the Dutch disease, the crowd-out of the manufacturing sector after a resource boom (Allcott and Keniston, 2018). In my case, changes in transportation infrastructure have the potential to generate improvements in the manufacturing sector. In a country in which industry linkages are such that the access to intermediate inputs has a major impact on the costs of the manufacturing sector, specific improvements in transportation can offset the crowd-out of the manufacturing sector caused by a commodity boom.

The rest of the paper is organized as follows. Section 2 describes the data and provides motivating facts. Section 3 presents the model. Section 4 describes how I take the model to data. Section 5 reports the results of my counterfactual exercises. Section 6 concludes.

1.2. Data and basic patterns

1.2.1. Data

This paper combines five datasets that allow me to measure domestic sectoral trade flows between Colombian departments, international trade flows between departments and the rest of the world by sector and port of exit/entry, input-output linkages, domestic trade costs, and international trade costs. My analysis focuses in four sectors (agriculture, mining, manufacturing and services) and considers data for 2013.

Customs data. I use a dataset created by the National Directorate of Taxes and Customs (DIAN, in Spanish) and the National Administrative Department of Statistics (the official statistical agency of Colombia, or DANE in Spanish) that contains all the shipments of exports and imports of Colombia. The data includes information such as harmonized system code, the department of origin/destination, and the city-port of exit/entry.[3]

Transportation and geography. I create a fully digitized road network that represents the primary highway system of Colombia,[4] based on physical and digital maps of the Ministry of Transportation and the National Institute of Roads (INVIAS). My main analysis focuses on roads, given that the share of total shipments (measured in tons) shipped via road is 73%, as of 2013 (ANIF, 2014).[5] For each highway segment, I have information on whether the road is paved, if it crosses a city, and whether the road is under public management or administered by a public-private partnership via the legal figure of *concesion*. Roads under the legal status of *concesion* are paved and tend to have better geographical and topographical characteristics than the rest of the roads.[6]

I estimate the travel times using Dijkstra's algorithm. I assign a speed of 30 km/hour for unpaved roads. The speeds for paved roads are 50 km/hour for paved roads in urban areas, 80 km/hour for paved highways outside urban centers, and 100 km/hour for paved roads under the legal figure of *concesion*. The speeds for paved and unpaved roads are

[3]I define a city-port as the location through which the products exit/enter the country. In the customs data, there is a total of 19 city-ports that are actively used for international shipments. The use of a city-port is based on the fact that goods could exit via a specific city, through different methods. For example, firms could use the seaport or the international airport located in Cartagena. In such cases, I do not differentiate by the method of transportation. Hence, in this example I would define Cartagena as a city-port of exit.

[4]Given that the transportation of goods mainly occurs via trucks, I do not consider the secondary road system (composed by roads administered by the Departments) nor the tertiary road system (managed by municipalities) and I focus exclusively in the primary road system. I do this because I do not have the status of the secondary or tertiary roads. Moreover, there are maps elaborated by the Ministry of Transportation, which contains graphical data about the annual flow of trucks by road. These maps show that most of the truck traffic use the primary road system. See IGAC (2005) for the most recent maps regarding truck flows across the country.

[5]The use of fluvial shipments is very limited, the railroad network is used exclusively for a specific route for the transportation of commodities, and the use of air cargo for domestic trade is relatively small (Duranton, 2015)

[6]Pachon and Ramirez (2006) explain that since the mid-90s, the Colombian government partially privatized some segments of the primary road system under the legal figure of public-private partnerships (concesiones, in Spanish). These roads were renovated/built by private companies, and the payments are split in two types: a direct government payment and the income generated by charging a fixed-fee to users of the highways).

like the ones used by Allen and Atkin (2016) for the Indian highway system, with the difference that I define different speeds for paved roads under *concesion*. I describe in the Appendix A why I consider the roads under *concesion* to be of higher quality, which leads me to assign them higher speed values.

Survey of cargo flows. I use the 2013 Survey of Origin/Destination of Cargo Transportation of the Ministry of Transportation to obtain proxies of domestic trade flows for the agricultural and manufacturing sectors. Specifically, I use the data on total weight cargo flows between different Colombian locations, measured in metric tons. Additionally, I use data regarding oil production and refining from the Ministry of Energy and Mines and the public oil company Ecopetrol, to generate domestic trade flows for the mining sector.

Input-output linkages. Data to calibrate the parameters of input-output linkages come from two sources: the World Input-Output Table of 2013 (Timmer, Dietzenbacher et al., 2015) and Colombia's input-output table produced by DANE for the year 2010.

1.2.2. Motivating facts

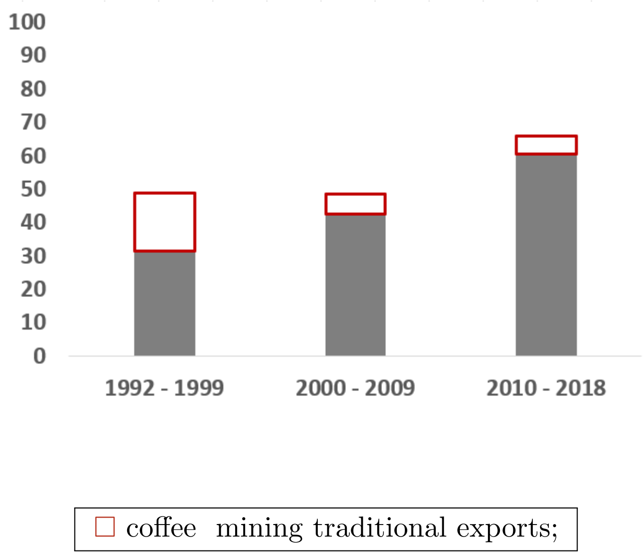
This section describes four empirical facts about Colombian departments that motivate the theoretical framework. First, Colombian exports are concentrated in a few goods, mostly mining ones. Second, the Colombian departments specialize in different sectors. Third, departments differ in their access to international markets, which generates differences in the international trade costs between departments and the rest of the world. Lastly, when the departments trade with the rest of the world, they do not use a single city-port to trade.

Fact 1, Colombian exports are concentrated in a few goods. Figure 1.1 plots the share of exports of *traditional products* as a fraction of total exports. This category was created by the Colombian government agencies for specific goods, given the historical concentration of exports in these products.[7] As figure 1.1 shows, during the past three decades, Colombia experienced an upward trend in the specialization of mining goods.

[7] This term is commonly used by government agencies such as the National Department of Planning or the statistical agency DANE. It groups the following products: coal, oil, coffee, and nickel-alloy.

Colombia was considered the standard case of an agricultural commodity-dependent nation by international agencies due to its dependence on coffee exports (FAO, 2002). More recently, an oil boom has reduced the share of coffee in the national exports. Recent official documents elaborated by the Colombian government highlight the dependence of the country on commodity exports (DNP, 2019).

Figure 1.1 Share of "traditional exports" according to Colombia's statistical agency DANE (%)



Notes: The bars show the average annual share of "traditional exports" with respect to total exports, for the period indicated in the x-axis. The source of the data is the official website of DANE.

Fact 2, Colombian departments specialize in different sectors. Using customs data from 2013, I build a Regional Index of Revealed Comparative Advantage (RCA) for every department. My objective is to show how a department specializes in a sector, relative to the specialization of Colombia in this same industry. The formula of this index is

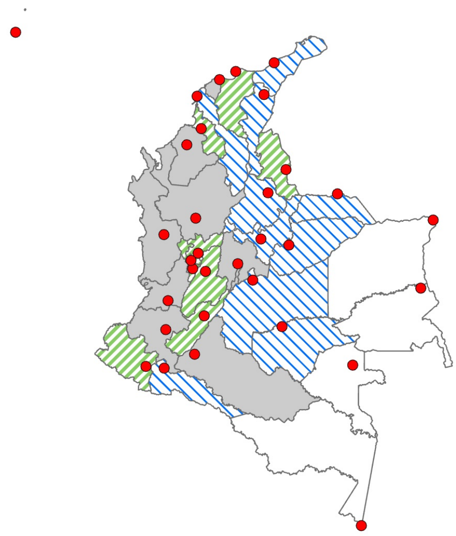
$$RCA_{s,d} = \left(\frac{Exports_{s,d}}{Total\ Exports_d}\right) / \left(\frac{Exports_{s,Colombia}}{Total\ Exports_{Colombia}}\right)$$

where s stands for a sector and d is a department. The index is the proportion of the exports of a department in sector s, divided by the proportion of Colombia's exports in industry s.

Intuitively, if the value of this ratio is high, a department is more specialized in sector s relative to the level of specialization of the entire Colombian economy in this industry. To obtain the Balassa Index of Revealed Comparative Advantage of every region (Balassa, 1966), the Regional index needs to be multiplied by the Balassa Index for Colombia. I use a regional index, instead of the Balassa index because I want to measure how every region is different than the Colombian economy, in its trade with the rest of the world.

After I obtain the values of the index, I select the sector in which every department shows the highest level of specialization. With this information, I construct figure 1.2 to provide evidence that there is variation in the sectoral specialization of Colombian regions.

Figure 1.2 Map indicating the sector with the strongest comparative advantage of every department (highest value of the Balassa Index)

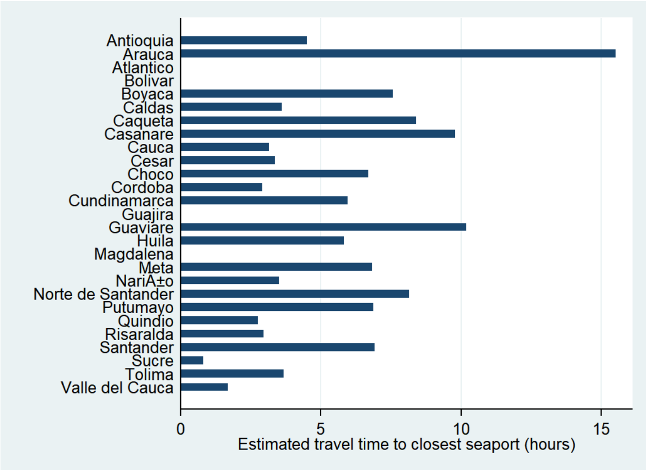


//// agriculture ; \\\\ mining; ■ manufacturing; • city-port

Notes: I do not consider the departments of Guainia, Leticia, San Andres y Providencia, Vaupes, and Vichada. Additionally, I merge Bogota with the department of Cundinamarca. See Appendix A for more details.

Fact 3, Colombian regions do not have uniform access to international markets. Colombian departments have heterogeneity in their access to global markets, given the existing geography of the country and the structure of the transportation network. To show this, Figure 1.3 displays the estimated travel times between the capitals of every department and the seaports of the country. (Given that 86% of exports and 70% of imports in 2013 exit or entered the country via seaports, figure 1.3 helps to illustrate the access to international markets of every Colombian department). The figure illustrates how some departments have immediate access to seaports, while for others it takes more than five hours to reach these ports.

Figure 1.3 Estimated travel times between the capital of the department and the closest seaport



Notes: I estimate the travel times between the capital of every department and the closest seaport using Dijkstra's algorithm, according to the speed values described in section 2.1. I do not consider the departments of San Andres y Providencia, Guainia, Leticia, Vichada and Vaupes. See Appendix A for details about this.

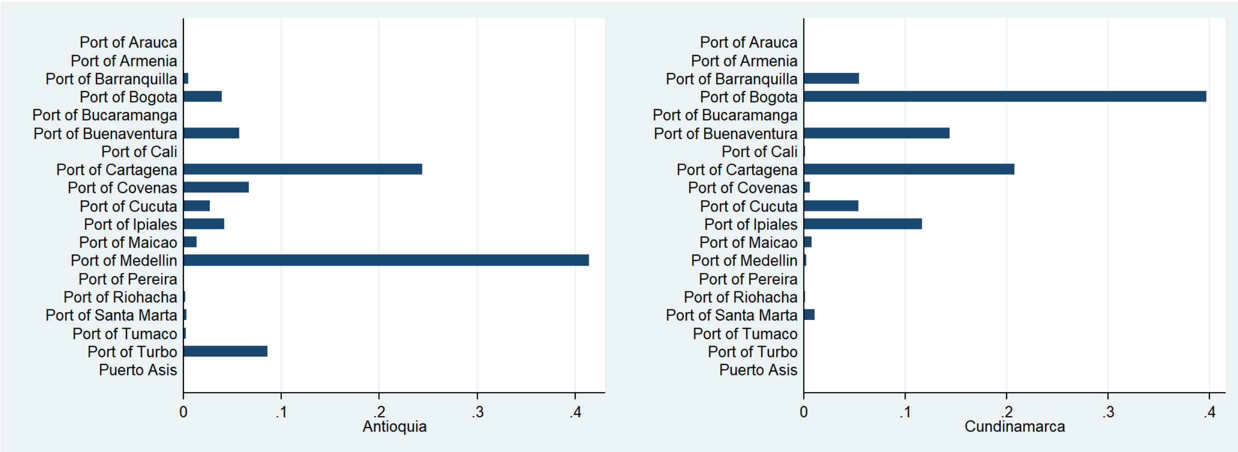
Fact 4, Colombian departments use multiple ports to trade with the rest of the world. Several departments have enough logistical infrastructure to trade with the rest of the world, such as airports, international land bridges, and seaports. In spite of this, most of the firms in the departments use different city-ports to trade with the global markets.

Figure 1.4 shows that the goods exported by the largest two departments of Colombia (Cundinamarca and Antioquia) are sent to other countries via different city-ports, even though Cundinamarca and Antioquia have large city-ports to serve international trade shipments.[8] The main explanation for this is that every city-port has logistical advantages for the shipment of specific goods, even within the same sector. For example, if I look at manufacturing goods, the seaport of Covenas is ideal for naphta products (a chemical manufacturing good), the airport of Bogota has excellent logistical conditions for the

[8]The city of Bogota located in the department of Cundinamarca posses the largest airport in the country, El Dorado International Airport, which has capacity to handle cargo shipments. The city of Medellin located in Antioquia has the Jose Maria Cordova International Airport, which also has infrastructure for the shipment of cargo.

shipment of textiles, while the seaport of Santa Marta has very good logistical capacity for handling steel and cement products.

Figure 1.4 Use of city-ports to export goods by the largest two Colombian departments (% of total department exports)



Notes: The vertical axis considers the 19 city-ports included in the customs data. For more details about the city-ports, see Appendix A.

1.3. Model

In this section, I describe my theoretical framework, define the equilibrium concept, provide an expression for a gravity equation, and explain how to translate changes in the road system into changes on the trade costs.

1.3.1. General framework

Geography. Consider an economy composed of Colombian departments and the rest of the world. These locations trade with each other. The departments are indexed by d and the rest of the world is indexed by RoW. The set of Colombian departments is $D = \{1, ..., \bar{d}\}$ and the set of all locations is $Z = \{1, ..., \bar{d}, RoW\}$. Each location is indexed by subscripts $n, j \in Z$. Trade between departments and the rest of the world require the use of city-ports ρ (see figure 1.5). There is a total of $\bar{\rho}$ city-ports. The set of city-ports is $\mathbb{P} = \{1, 2, ..., \bar{\rho}\}$.

I define an international shipping route as an ordered pair that consists of a department d and a city-port ρ. An *export route* consists of an ordered pair department, city-port $r_x = (d, \rho)$. There is a total of $\bar{d}\bar{\rho}$ export routes. The set of export routes is $R_x = D \times \mathbb{P}$. The subset of export routes for a department d is defined as $R_{x,d} = \{(d, \rho) : \rho \in \mathbb{P}\}$. An *import route* consists of an ordered pair city port-department $r_m = (\rho, d)$. There are $\bar{d}\bar{\rho}$ import routes. The set of import routes is $R_m = \{\mathbb{P} \times D\}$. The subset of import routes for a department d is defined as $R_{m,d} = \{(\rho, d) : d \in D\}$

Goods. There are two types of goods, intermediates and composite goods. There are four sectors in the economy: agriculture (a), mining (m), manufacturing (i) and services (z). Sectors are indexed by $k \in \{a, m, i, z\}$. Intermediate good firms in location n and sector k produce intermediate good. Firms that produce composite goods buy from suppliers across different locations and produce an aggregated composite using a Dixit-Stiglitz aggregator. The market structure in all sectors is perfect competition.

Trade costs between departments and the rest of the world. International trade between a department, d, and the rest of the world, RoW, require specialized traders, as in Allen and Arkolakis (2019). There is a continuum of specialized traders $\iota \in [0, 1]$. Traders choose among all the shipping routes when they export or import goods.[9] These traders face capacity constraints when moving goods internationally.

Figure 1.5 helps to understand the concept of international shipping routes. For example, when department 1 trades with the rest of the world, *Route 1A* can be used, which implies that the shipment of goods occurs via *Port A*. Or *Route 1B* can be used, therefore, the *Port B* will be chosen for the shipment of goods. A similar logic occurs when department 2 trades with the rest of the world.

Every specialized trader faces a productivity shock that is specific to the international shipping route and to every sector k. This implies that the cost of a specialized trader ι when it uses an international shipping route r_t is $\tau_{r_t,k}/z_{r_t,k}(\iota)$. I define the international shipping cost for trader ι as the lowest international shipping cost across different routes,

[9]Intuitively, firms choose logistical companies to ship goods between a department and the rest of the world (e.g. Fedex, UPS, McLane Company, JR Freight, etc.)

when the trader ships a good between department d and RoW, that is

$$\tau(\iota) = \underset{r_t}{min} \frac{\tau_{r_t,k}}{z_{r_t,k}(\iota)} \text{ for } t \in \{x, m\} \tag{1.1}$$

where τ_{r_t} is the shipping cost along route r_t for goods of sector k, $z_{r_t,k}(\iota)$ is the productivity draw for a specific international shipping route r_t to transport goods of sector k, and subscript t defines whether the shipping route is used to export or import goods. This productivity draw follows a Frechet distribution with parameters $(A_{r_t,k}, \theta_k)$. The Frechet parameter $A_{r_t,k}$ is the scale parameter of the Frechet distribution. The shape parameter θ_k represents the heterogeneity of productivities of city-ports regarding the transportation of sector-k goods. The higher the value of θ_k, the lower the heterogeneity in the productivities of city-ports. Thus, high values of θ_k imply that traders tend to use the same city-port to move goods between departments and the rest of the world.

When agents buy exported or imported goods, they are randomly assigned with specialized traders. Thus, the iceberg trade cost between a department d and the rest of the world RoW is the expected trade cost across the continuum of traders, as in Allen and Arkolakis (2019).

$$\tau_{dRoW,k} \equiv E[\tau(\iota)] = E\left[\underset{r_t,k}{min} \frac{\tau_{r_t}}{z_{r_t,k}(\iota)}\right] \tag{1.2}$$

Using the properties of Frechet distribution, the expression for the icerberg trade cost between any department d and the rest of the world becomes

$$\tau_{dRoW,k} = \Phi_x^{-\frac{1}{\theta}} \Gamma\left(\frac{1+\theta_k}{\theta_k}\right) \text{ where } \Phi_k = \sum_{r_t} A_{r_t} \tau_{r_t}^{-\theta_k} \tag{1.3}$$

where Γ is the gamma function.

International shipping costs. Following Duranton, Morrow, and Turner (2014), I define the international shipping cost of route $r_t = (d, \rho)$ as $\tau_{r_t} \equiv \tau_\rho \tau_{d\rho} \tau_d$. This implies that the international shipping cost of a route depends on logistical characteristics of department d, denoted by τ_d, the logistical capacity of the port ρ, represented by τ_ρ, and the connectivity between department d and port ρ, expressed as $\tau_{\rho d}$. The latter is a function of the travel times between d and ρ, T_{dp}, therefore $\tau_{d\rho} = f(T_{d\rho})$.[10]

Trade costs between departments in Colombia. There are standard iceberg trade costs for every sector. I denote the trade costs between department $d_1 \in D$ and department $d_2 \in D$ for sector-k goods as $\tau_{d_1 d_2,k}$. Icerberg trade costs between departments are a function of travel times along the least cost route that connects these departments ($T_{d_1 d_2}$), that is $\tau_{d_1 d_2} = f(T_{d_1 d_2})$. Notice that in figure 1.5 there is only one route to move goods between department 1 and 2.

Domestic traders are homogeneous, hence they always choose the same optimal road when sending goods from d_1 to d_2. Implicitly, this implies that all the trade flows are shipped through the least cost road between d_1 and d_2. If I consider the existence of traders for domestic trade, this assumption can be interpreted as having a very high value for the shape parameter θ that represents the heterogeneity in the use of roads across two locations within Colombia. The assumption is consistent with Allen and Arkolakis (2019), who find that domestic traders moving goods across two cities within a country tend to choose the same least cost road.

Preferences. Consumers' preferences are represented by a Cobb-Douglas utility function given by

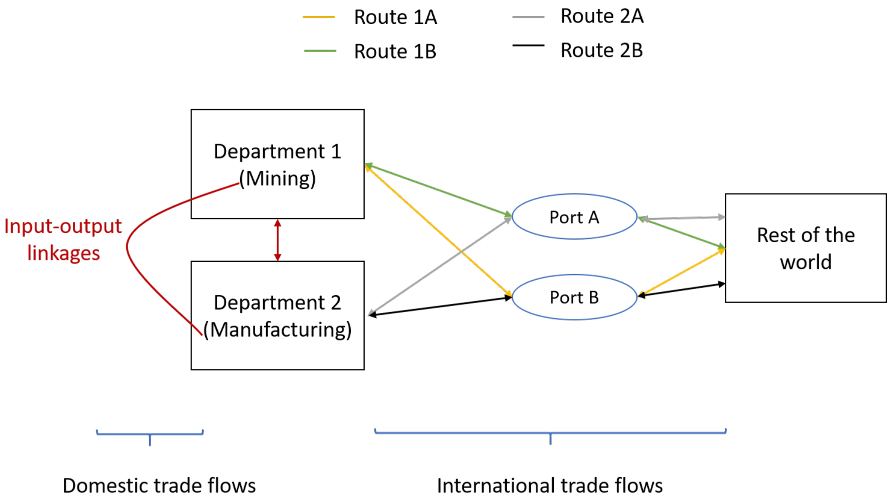
$$U_j = \prod_{k=1}^{K} (C_j^k)^{\alpha_j^k} \text{ , with } \sum_{k=1}^{K} \alpha_j^k = 1 \tag{1.4}$$

where α_j^k is the share of sector k in final demand and C_j^k is the level of consumption of the composite good. The income of households is denoted by I_n. Households' income are the sum of payments to labor and transfers, that is $I_n = w_n L_n + D_n$. The transfers are equal to deficits as in Dekle, Eaton and Kortum (2008).

Labor supply. Agents live in location $n \in Z$ and supply one unit of labor. There are L_n workers in location n. There is perfect labor mobility across sectors, but no labor mobility across locations (this implies no labor mobility across Colombian departments).

[10] The assumptions have implications about the symmetry of shipping costs of export and import routes $\tau_{r_x} = \tau_{r_m} = \tau_d \tau_{dp} \tau_\rho \ , \ if \ r_x = (d, \rho) \ and \ r_m = (\rho, d)$

Figure 1.5 Economic environment



1.3.2. Production

Production of intermediates. The production of intermediate goods requires labor and composite goods from all sectors. Technology has constant returns to scale and it is defined by

$$q_{n,k} = A_{n,k} l_{n,k}^{\beta_n^{l,k}} \Big[\prod_{s\{a,m,i,z\}} m_{s,k}^{\beta_n^{s,k}} \Big] \tag{1.5}$$

where $\beta_n^{l,k} + \sum_s \beta_n^{s,k} = 1 \; \forall \; n$. I denote by $m_{s,k}$ the amount of composite good of sector s used in the production of sector k, $\beta_n^{s,k}$ is the parameter that defines the share of composite goods from sector s used in the production of intermediates for sector k goods, $\beta_n^{l,k}$ is the share of value added of sector k, $A_{z,k}$ is the productivity of sector k, l_n^k is the amount of labor necessary for the production of good of sector k in city n,

Firms price at unit cost $\frac{c_{n,k}}{A_{n,k}}$, where $c_{n,k}$ is the unit cost of an input bundle. This can be expressed as

$$c_{n,k} = \phi_{n,k} (w_n)^{\beta_n^{l,k}} \prod_s (P_{n,s})^{\beta_n^{s,k}} \tag{1.6}$$

where $\phi_{n,k} \equiv (\beta_n^{l,k})^{-\beta_n^{l,k}} (\beta_n^{a,k})^{-\beta_n^{a,k}} (\beta_n^{i,k})^{-\beta_n^{i,k}} (\beta_n^{m,k})^{-\beta_n^{m,k}} (\beta_n^{z,k})^{-\beta_n^{z,k}}$ is a constant, and $P_{n,s}$ is the price of a composite intermediate good from sector s in location n. The cost function captures the input-output linkages between industries: if the price of the composite good in one industry changes, it will affect the unit cost of the rest of the sectors.

Production of composite goods. Firms that produce composite goods in location n for sector k purchase the intermediate goods from suppliers across different locations. The production technology of composite goods uses a Dixit-Stiglitz aggregator:

$$Q_{n,k} = \Big[\sum_j (q^c_{jn,k})^{\frac{\sigma_k - 1}{\sigma_k}} \Big]^{\frac{\sigma_k}{\sigma_k - 1}} \tag{1.7}$$

where $Q_{n,k}$ is the number of units that the firms supply, σ_k is the elasticity of substitution between intermediates of sector k and $q^c_{jn,k}$ is the demand of intermediate good of sector k by city n produced in city j.

Prices. Given the existence of perfect competition, the price of a good of sector k consumed by location n and produced in j considers the unit cost and the trade costs between locations, that is

$$p_{jn,k} = \frac{c_{j,k} \tau_{jn,k}}{A_{j,k}} \tag{1.8}$$

using this expression, I derive the price of the composite good of sector k in location n

$$P_{n,k} = \Big[\sum_j p_{jn,k}^{1-\sigma_k} \Big]^{\frac{1}{1-\sigma_k}} = \Big[\sum_j \Big(\frac{\tau_{jn} c_{j,k}}{A_{j,k}} \Big)^{1-\sigma_k} \Big]^{\frac{1}{1-\sigma_k}} \tag{1.9}$$

where the second equality comes from using (8). Using the previous prices of sector k, I can obtain the price index of location n:

$$P_n = \prod_k \Big(\frac{P_{n,k}}{\alpha_{n,k}} \Big)^{\alpha_{n,k}} \tag{1.10}$$

1.3.3. Trade flows and expenditure shares

Solving the optimization problem of the firms that produce the composite good, I obtain an expression for the demand of intermediate good in sector k, denoted by $q^c_{jn,k}$. Combining it with the price of intermediate good $p_{jn,k}$ and aggregating, I derive an expression for the total expenditure by location n on goods from sector k produced in location j

$$X_{jn,k} = \Big(\frac{\tau_{jn}c_{j,k}}{A_{j,k}}\Big)^{1-\sigma_k} Q_{n,k} P_{n,k}^{\sigma_k-1} \tag{1.11}$$

Following Anderson and van Wincoop (2003), the trade flows equation can also be expressed as

$$X_{jn,k} = (\tau_{jn})^{1-\sigma_k}\Big(\frac{Y_{j,k}}{\Pi_{j,k}^{1-\sigma_k}}\Big) Q_{n,k} P_{n,k}^{\sigma-1} \tag{1.12}$$

where $\Pi_{j,k}^{1-\sigma_k} \equiv \sum_m \tau_{jm}^{1-\sigma_k} X_{m,k} P_{m,k}^{\sigma_k-1}$. The term $X_{m,k}$ is the total expenditure of location m in goods of sector k. Finally, let $\lambda_{jn,k}$ be the fraction of expenditure of j in sector-k goods produced by location n:

$$\lambda_{jn,k} \equiv \frac{X_{jn,k}}{\sum_l X_{ln,k}} = \Big(\frac{\tau_{jn}c_{j,k}}{A_{j,k}}\Big)^{1-\sigma_k} (P_{n,k})^{\sigma_k-1} \tag{1.13}$$

1.3.4. Total expenditure and trade balance

The total expenditure of location n in sector-k goods $X_{n,k}$ is composed by the expenditure by firms on intermediates (that depends on total exports of location n) and the households' expenditure (which is a constant fraction $\alpha_{n,k}$ of the total income):

$$X_{n,s} = \sum_k \beta_n^{s,k} \sum_j X_{j,k}\lambda_{nj,k} + \alpha_{n,s} I_n \tag{1.14}$$

where I_n denotes the total income of sector n, composed by labor income and transfers. The total income in location n is $I_n = w_n L_n + D_n$, where D_n is the total deficit of n.

The total trade deficits sum up to zero across all locations ($\sum_n D_n = 0$) and the total trade deficits are the sum of sectoral trade deficits, $D_n = \sum_k D_{n,k}$. A sectoral trade deficit $D_{n,k}$ is defined as $D_{n,k} = M_{n,k} - E_{n,k}$ where $M_{n,k} = \sum_j X_{n,k}\lambda_{jn,k}$ represents the total imports of country n of sector-k goods and $E_{n,k} = \sum_j X_{j,k}\lambda_{nj,k}$ is the total exports of n of sector-k goods. I consider total trade deficits as exogenous, but the sectoral trade deficits are endogenous, as in Caliendo and Parro (2015).

Considering the definition of total trade deficit for any location n, I can express the trade balance equation as

$$\sum_k \sum_j X_{j,k}\lambda_{nj,k} = \sum_k \sum_j X_{n,k}\lambda_{jn,k} - D_n \tag{1.15}$$

Labor market clearing. By aggregating the total expenditure of location n in sector k, equation (14), across all sectors and combining it with the trade balance equation (15), I get an expression for the labor market clearing (see Appendix B).

$$w_n L_n = \sum_k \beta_n^{l,k} \sum_j X_{nj,k} = \sum_k \beta_n^{l,k} \sum_j X_{j,k}\lambda_{nj,k} \tag{1.16}$$

1.3.5. Equilibrium

In this section, I define the world equilibrium. Then, I describe the equilibrium in changes, which requires fewer parameters than the original equilibrium. By doing this, I simplify the estimation procedure.

Equilibrium in levels

Definition 1. World equilibrium in levels. *The equilibrium is a set of wages $\{w_{n,k}\}_{n\in Z,k\in\{a,m,i,z\}}$, prices $\{P_{n,k}\}_{n\in R,k\in\{a,m,i,z\}}$, and labor allocations $\{L_{n,k}\}_{n\in Z,k\in\{a,m,i\}}$ for all locations $n \in Z$ under the assumption of perfect labor mobility across sectors and immobile labor across locations that solve equations (6), (9), (13), (14) and (15).*

Equilibrium in changes

Solving the previous equilibrium requires the knowledge of many parameters that are difficult to estimate, such as the sectoral productivities $\{A_{j,k}\}$. An option to reduce the number of parameters needed to calibrate the model, is to express the equilibrium in changes.

Following Dekle, Eaton and Kortum (2008), let x' be the value of any variable in the new steady state and define the change in the value of variables between the old and the new equilibrium as $\hat{x} = x'/x$. Thus, I obtain an expression for any variable in the new equilibrium as $x' = \hat{x}x$. The following definition, considers the original equilibrium in terms of changes. This is similar to Caliendo and Parro (2015).

Definition 2: Equilibrium in terms of changes. *Let* $(\mathbf{w}, \mathbf{P})$ *be an equilibrium under trade costs* $\{\tau_{\mathbf{jn}}\}_{\mathbf{j,n \in R}}$. *Consider a different equilibrium* $(\mathbf{w}', \mathbf{P}')$ *under trade costs* $\{\tau'_{\mathbf{jn}}\}_{\mathbf{j,n \in R}}$. *Let* $(\hat{w}, \hat{P})$ *be an equilibrium under trade costs* $\{\tau'_{\mathbf{jn}}\}_{\mathbf{j,n \in R}}$ *relative to* $\{\tau_{\mathbf{jn}}\}_{\mathbf{j,n \in R}}$, *where variable* $\hat{x}$ *represents relative changes, that is* $\hat{x} = \frac{x'}{x}$. *Then, the equilibrium conditions (6), (9), (13), (14) and (15) can be expressed in relative changes:*

i *Good market clearing condition*

$$\hat{c}_{n,k} = (\hat{w}_n)^{\beta_n^{lk}} \prod_{s \in \{a,m,i,z\}} (\hat{P}_{ns})^{\beta_n^{sk}} \tag{1.17}$$

ii *Expenditure shares*

$$\hat{\lambda}_{jn,k} = (\hat{\tau}_{jn,k})^{1-\sigma_k} (\hat{c}_{j,k})^{1-\sigma_k} (\hat{P}_{n,k})^{\sigma_k - 1} \tag{1.18}$$

iii *Prices*

$$\hat{P}_{nk} = \Big[\sum_j (\hat{\tau}_{jn} \hat{c}_{j,k})^{1-\sigma_k} \lambda_{jn,k} \Big]^{\frac{1}{1-\sigma_k}} \tag{1.19}$$

iv *Total expenditure*

$$X'_{n,s} = \sum_k \beta_n^{s,k} \sum_j X'_{j,k} \lambda'_{nj,k} + \alpha_{n,s} I'_n$$

$$X'_{n,s} = \sum_k \beta_n^{s,k} \sum_j X'_{j,k} \hat{\lambda}_{nj,k} \lambda_{nj,k} + \alpha_{n,s} [\hat{w}_n w_n L_n + D'_n] \tag{1.20}$$

v *Trade balance*

$$\sum_k \sum_j X'_{j,k} \lambda'_{nj,k} = \sum_k \sum_j X'_{n,k} \lambda'_{jn,k} - D'_n$$

$$\sum_k \sum_j X'_{j,k} \hat{\lambda}_{nj,k} \lambda_{nj,k} = \sum_k \sum_j X'_{n,k} \hat{\lambda}_{jn,k} \lambda_{jn,k} - D'_n \tag{1.21}$$

1.3.6. Department-port gravity equation

I generate an expression for international trade flows between department d and the rest of the world, RoW, that use a specific city-port ρ (or specific international shipping route r_t). For the case of those trade flows between the rest of the world and the departments, equation (12) becomes a different expression. This is necessary, given I need to include the role of the specialized traders on the international trade flows. To do this, I obtain the share of exports/imports that use route r_t and combine it with equation (3), which defines the relationship between international shipping costs and trade costs, to generate a department-port gravity equation.

Shares of international shipping routes. Using the properties of the Frechet distribution, it is possible to obtain an expression for the shares of trade flows that are shipped via a specific international shipping route r_t for $t \in \{m, x\}$. Define $G_{r_t}(c)$ as the probability that the *international shipping cost* of a good sent via route r_t is lower than c.

$$G_{r_t,k}(c) \equiv Pr\Big[\frac{\tau_{r_t}}{z_{r_t,k}(\iota)} \leq c\Big]$$

$$G_{r_t,k}(c) = 1 - exp[-A_{r_t}(\tau_{r_t})^{-\theta_k} c^{\theta_k}] \tag{1.22}$$

Let $G_{t,k}(c)$ be the probability that a good shipped via route r_t has an *observed cost* lower than c. This probability is expressed as

$$G_{t,k}(c) \equiv Pr\{\tau_s(\iota) \leq c\} = Pr\Big[\min_{r_t} \frac{\tau_{r_t,k}}{z_{r_t,k}(\iota)} \leq c\Big]$$

$$G_{t,k}(c) = 1 - exp[-c^{\theta} \Phi_{t,k}], \quad \text{where} \quad \Phi_k = \sum_{r_t} A_{r_t} \tau_{r_t}^{-\theta_k} \tag{1.23}$$

Finally, define π_{r_t} as the probability that a good is shipped via route r_t as

$$\pi_{r_t,k} = Pr\{\tau_{r_t,k}(\iota) \leq \underset{v_t \in R_{t,d} \setminus r_t}{min} \tau_{v_t,k}(\iota)\}$$

$$\pi_{r_t,k} = \frac{A_{r_t}\tau_{r_t}^{-\theta_k}}{\Phi_{t,k}} \tag{1.24}$$

Similar to Eaton and Kortum (2002), I can show that the distribution of international shipping costs is the same, no matter which route is used (see Appendix B). This implies that $\pi_{r_t,k}$ also represents the share of the value of exports/imports between a department d and RoW, sent via route r_t.

Trade flows between department and rest of the world via a city-port. I obtain an expression for the trade flows between departments and the rest of the world shipped via a specific route r_t. Consider as example, the export flows that use route $r_t = (d, p)$:

$$X_{dRoW,k,r_t} = X_{dRoW,k}\pi_{r_t,k}$$

$$X_{dRoW,k,r_t} = (\tau_{dRoW,k})^{1-\sigma_k}\left(\frac{Y_{RoW,k}}{\Pi_{RoW,k}^{1-\sigma_k}}\right)Q_{d,k}P_{d,k}^{\sigma-1}\pi_{r_t,k}$$

Inserting (3) and (24) into the expression for trade flows between any department d to RoW, that are sent via route $r_t = (d, p)$, I get:

$$X_{dRoW,k,dp} = \left[\Phi_k^{-\frac{1}{\theta}}\Gamma\left(\frac{1+\theta}{\theta}\right)\right]^{1-\sigma_k}\left(\frac{Y_{RoW,k}}{\Pi_{RoW,k}^{1-\sigma_k}}\right)Q_{d,k}P_{d,k}^{\sigma-1}\left[\frac{A_dA_\rho(\tau_d\tau_p\tau_{dp})^{-\theta_k}}{\Phi_{t,k}}\right] \tag{1.25}$$

To obtain the previous result, I assume that $A_{r_t} = A_{dp} = A_dA_\rho$. This implies that the scale parameter of the Frechet distribution, which governs the behavior of the productivities of the shipping routes, depends on a productivity transportation factor related to the department, and another productivity transportation factor related to the ports. The assumption is economically intuitive. To see this, notice that if any of these factors increases, then the international trade costs between departments and the rest of the world fall (see equation 3), and the probability that the route $r_t = (d, \rho)$ is used also increases (see equation 24).

For the international shipping costs, I use the expression $\tau_{r_t} = \tau_p \tau_{dp} \tau_p$. A similar expression can be obtained for imports using a particular international shipping route. Notice that the assumption regarding the productivity term for the international shipping routes implies symmetric trade costs.

There are two characteristics of the international shipping costs τ_{dp} that matter for the theoretical framework. First, they affect the share of trade flows X_{dRoW} and X_{RoWd} that are traded via port ρ through international shipping routes $r_x = (d, \rho)$ and $r_m = (\rho, d)$, respectively, through the term π_{r_t}. Second, the international shipping costs affect the trade costs between department d and the rest of the world, $\tau_{d\ RoW}$. Such effects are economically intuitive. Consider that τ_{dp} depends on the infrastructure that connect d and ρ. If an infrastructure project reduces the road distance between d and ρ, then port ρ will be used more often ($\uparrow \pi_{r_t}$), and the department d will better connected to the global markets ($\downarrow \tau_{dRoW}$).

1.3.7. Estimation of changes in trade costs due to new infrastructure projects

I can use the *equilibrium in changes* previously defined in section 3.5 only if I take as given a specific change in the vector of trade costs, $\hat{\tau}$. The objective of this paper is to evaluate how a new road infrastructure project change the national comparative advantage. Hence, I need to define how improvements in the Colombian road network lead to changes in trade costs. To facilitate the comprehension of this process, figure 1.6 illustrates how new infrastructure projects translate into changes in trade costs.

Estimation of the change in trade costs between departments and the rest of the world. Consider a large infrastructure project that changes the travel times across all international shipping routes from $\{T_{r_t}\}$ to $\{T'_{r_t}\}$. If the function between trade costs and travel times is known, $\tau = f(T)$, then it is possible to obtain both the old and the new international shipping costs along all routes, τ_{r_t} and τ'_{r_t}, respectively. I use the function $\tau_{r_t} = exp(\beta_{time} T_{r_t})$, which is a standard assumption in international trade and economic geography models. I discuss with detail how to obtain the value of the parameter β_{time} in section 4.

Using the exact algebra method of Dekle, Eaton and Kortum (2009) with the transportation model equations (3) and (24), I can obtain the change in shares of trade flows between d and RoW that use international shipping route r_t

$$\hat{\pi}_{r_t,k} = \frac{(\hat{\tau}_{r_t})^{-\theta_k}}{\sum\limits_{v_t \in R_t} \pi_{v_t,k}(\hat{\tau}_{v_t})^{-\theta_k}} \tag{1.26}$$

and the change in trade costs between department d and RoW is expressed as

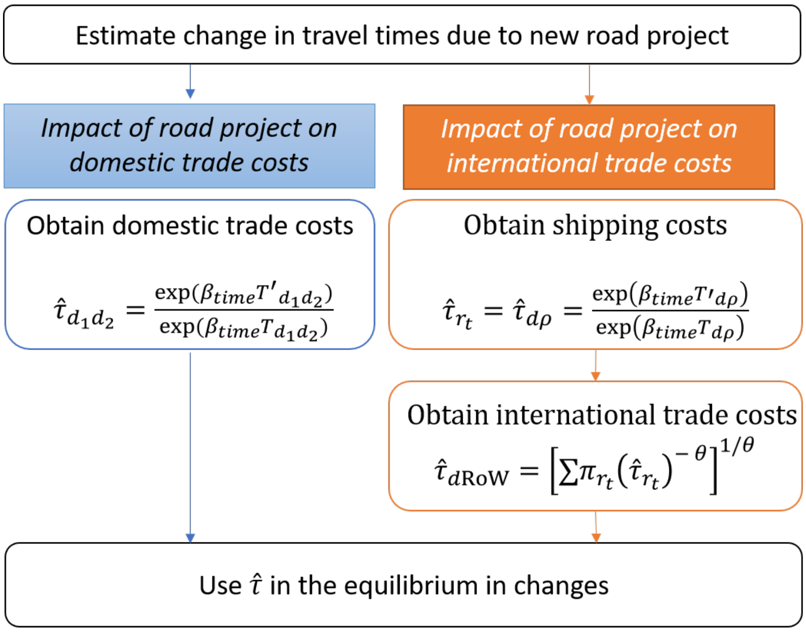
$$\hat{\tau}_{dRoW,k} = \Big[\sum_{r_t \in R_t} \pi_{r_t,k}(\hat{\tau}_{r_t})^{-\theta_k}\Big]^{-\frac{1}{\theta_k}} \tag{1.27}$$

where π_{dRoW,r_t} is the share of exports of department d to the rest of the world that use the route r_t. I can estimate this share using customs administrative data.[11].

Estimation of the changes in trade costs between departments. I obtain the travel times before the infrastructure project is built,$\{T_{d_1d_2}\}_{d_1,d_2 \in D}$, and after the highway is completed, $\{T'_{d_1d_2}\}_{d_1,d_2 \in D}$. Then, I can get both the old and the new trade costs between departments ($\tau_{d_1d_2}$ and $\tau'_{d_1d_2}$, respectively) using directly the function $\tau_{d_1d_2} = f(T_{d_1d_2}) = exp(\beta_{time}T_{d_1d_2})$. I do this because I assume there is no heterogeneity in the use of shipping routes between any two departments. Once I obtain the old and the new trade costs for the domestic trade model, I can calculate directly the change in trade costs for trade flows across departments, $\hat{\tau}_{d_1d_2} = \frac{\tau'_{d_1d_2}}{\tau_{d_1d_2}}$.

[11]The shares of the export flows transported through a specific route might not necessarily be the same as the shares of imports shipped through this route (i.e. $\pi_{dRoW,r_t} \neq \pi_{RoWd,r_t}$). Hence, to make the counterfactual consistent with symmetric trade costs, I estimate the change in trade costs between any department and the rest of the world using the shares of total trade flows.

Figure 1.6 Steps to obtain changes in trade costs



1.4. Taking the model to data

1.4.1. Parameters of the Armington model

Data sources to calibrate the model. I use the following datasets (i) customs data with records about individual export and import shipments, with information about the port of entry/exit, (ii) the World IO database (WIOD), (iii) the input-output table from the Colombian statistical agency for 2010, (iv) the 2013 Transportation Survey of Origin/Destination elaborated by the Colombian Ministry of Transportation, [12] (v) crude oil production data and refinery capacity, and (vi) the Economic Accounts produced by DANE to obtain variables such as value-added and gross output at a sectoral level . Appendix A provides more details.

Production and consumption parameters. I use the same value for the elasticity of substitution for all sectors, $\sigma_k = 6\ \forall k$. I estimate the share of value added for the rest of the world and the departments using $\beta_n^{l,k} = (VA_k)/Y_k$, where VA_k is value added of sector k and Y_k is the gross production. Given the lack of input-output tables for Colombian

[12]This data was used to generate a proxy of the domestic trade flows of agriculture and manufacturing. Unfortunately, Colombia does not have a detailed Commodity Flow Survey like the United States that allows researchers to estimate good measures of domestic trade flows

departments, I assume the same value for this parameter for all departments. I estimated the share of sector s in the production of sector k using $\beta_n^{s,k} = (1 - \beta_n^{l,k}) \frac{C_{intermediate,k,s}}{C_{intermediate,k,total}}$, where $C_{intermediate,k,s}$ is the intermediate consumption of sector k in goods from sector s, and $C_{intermediate,k,total}$ is the total intermediate consumption of sector k. I assume identical values of these parameters for all departments. Lastly, I estimate the share of final consumption in sector k with data from the input-output tables, using the formula $\alpha_{n,k} = C_{k,final,total}/C_{final,total}$, where $C_{k,final,total}$ is the final consumption in sector k and $C_{final,total}$ is the level of total final consumption.

Trade deficits and expenditure shares: agriculture and manufacturing. My estimation of trade deficits is limited by the information of transportation survey data that serves as proxy for domestic trade flows, for the agriculture and manufacturing sectors. The trade deficit of any department d can be considered as $D_{d,Total} = D_{d,Domestic\ trade} + D_{d,International\ trade}$. Unfortunately, I cannot obtain direct estimates of domestic trade flows using the cargo transportation survey. Hence, I assume that for Colombian departments, the deficit generated from the domestic trade is very small relative to the deficit the international trade. Hence, my deficit estimations exclusively consider the customs administrative data.

Expenditure shares: agriculture and manufacturing. For the case of the expenditure shares, $\lambda_{nj,k}$, table 1 illustrates the construction of the shares. I obtain the share of expenditures of Colombia on its goods, denoted by $\gamma_{Col,Col}$, using Colombia's input-output table, the share of expenditures of rest of the world on its goods, represented by $\gamma_{RoW\ RoW}$, using WIOD tables and the customs administrative dataset. Besides, using the customs data, I obtain the share of Colombian exports for every department, expressed as γ_{dRoW}, and the department share of national imports, characterized by γ_{RoWd}.

To obtain data on domestic trade flows, I rely on the transportation survey elaborated by the Ministry of Transportation for 2013. I assume this survey exclusively reflects patterns of domestic trade. I denote $\mu_{d_1 d_2}$ as the shares of expenditures of a department d_2 in goods from d_1, exclusively considering domestic trade flows. Notice these are not the shares from the Armington model $\lambda_{d_1 d_2}$ for $d_1, d_2 \in D$, because such shares consider both domestic and international trade. Unfortunately, I am not able to obtain values for the share of expenditures of departments on their goods for the case of domestic trade flows, μ_{dd}, therefore, I run my simulations under different values for this parameter ($\mu_{dd} = 0.3, 0.6$).

Table 1.1 Construction of the matrix of expenditure shares

Exporter ↓ Importer →	**RoW**	$\mathbf{d_1}$	...	...	$\mathbf{d_D}$
RoW	$\lambda_{RoW\,RoW} = \gamma_{RoW\ \ RoW}$	$\lambda_{RoW\,d_1} = (1-\gamma_{ColCol})\gamma_{RoW\,d_1}$			
$\mathbf{d_1}$	$\lambda_{d_1 RoW} = (1-\gamma_{RoWRoW})\gamma_{d_1 RoW}$	$\lambda_{d_1 d_1} = \mu_{d_1 d_1}\gamma_{ColCol}$			
...					
...					
$\mathbf{d_D}$	$\lambda_{d_D RoW} = (1-\gamma_{RoWRoW})$				$\lambda_{d_D d_D} = \mu_{d_D d_D}\gamma_{ColCol}$

Notes: μ_{ij} represents shares exclusively considering domestic trade flows between locations i and j, while γ_{mn} represents international trade flows between m and n

Trade deficits and expenditure shares: mining. I build the international and domestic expenditure shares of mining under the assumption that domestic trade flows are exclusively for crude oil between departments with oil fields and those with refineries,[13] while international trade flows include oil, coal, and minerals. I assume that those departments that are oil producers ship crude oil to the five refineries located in the departments of Bolivar, Santander, Casanare, Putumayo, and Meta.

Given that Colombia is a crude oil exporter, I presume that refineries only use crude oil produced domestically. To build these domestic trade flows, I infer that departments with refineries consume all the crude oil they produce, and if there is remaining capacity, they will import crude oil from other departments. The size of such domestic imports from each department is proportional to their oil production. This assumption allows me to obtain domestic trade flows for the mining sector. Additionally, I used the customs data to obtain international trade flows of the mining sector between departments and the rest of the world.[14]

[13] 86 % of the gross domestic output of the mining sector is coal and crude oil. According to the Energy International Agency, Colombia exports most of its coal production. Hence, I assume that the domestic trade flows consisted mostly of crude oil from departments with oil fields to departments with oil refineries

[14] I could build a more precise measure of mining domestic trade flows using pipelines information. Unfortunately, I do not have accurate geospatial data about pipeline location and capacity.

1.4.2. Solving the model

I solve the model using the algorithm of Caliendo and Parro (2015). I make two adjustments: I do not need to consider how tariffs affect the expenditure function, and my measure of welfare does not need to consider tariff revenue.

1.4.3. Parameters of the transportation model

The department-port gravity equation (25) does not allow me to estimate the parameters that determine the dispersion of productivity of the shipping routes by sector, θ_k. To see this, consider the standard assumptions in international trade and economic geography models regarding the relationship between trade costs and travel time.

$$\tau_{d\rho} = exp(\beta_{time} T_{d\rho}) \tag{1.28}$$

where $T_{d\rho}$ is the travel time between department d and ρ and β_{time} is the parameter that defines the relationship between the shipping costs of a route $r_t = (d, \rho)$ and the travel time between department d and city-port ρ. By inserting this expression in the department-port gravity equation (25), and taking logs I obtain

$$ln(X_{dRoW,k,d\rho}) = \alpha + \alpha_{d,exporter} + \alpha_{d,importer} + \alpha_{RoW,exporter} + \alpha_{RoW,importer} \tag{1.29}$$

$$+\alpha_{\rho} - \mu_{t,k}(T_{d\rho}) + \epsilon_{d\rho}$$

where $\mu_{t,k} = \theta_k \beta_{time}$, and T_{dp} is the travel time between department d and city-port ρ. Using this structural regression, I get an estimate of $\mu_{t,k}$. Given that I cannot estimate separately the parameters β_t and θ_k, I use the value of β_{time} from previous literature. Specifically, I use the value reported by Allen and Arkolakis (2019). I elaborate about the value of this parameter in section 4.5.

1.4.4. Estimation of gravity equation

Although it is possible to use OLS to estimate $\mu_{t,k}$ using (29), there are concerns about the presence of endogeneity given the existence of unobservables correlated with both the travel time between a department d and city-port ρ and the international trade flows between such pair, $X_{dRoW,k,d\rho}$. Consider that $\epsilon_{d\rho}$ represents a bilateral cost/demand shifter of the international trade flows using the route $r_t = (d, \rho)$. The main source of endogeneity is the fact that the Colombian national government could target the pair department city-port, (d, ρ), through infrastructure policies that affect both the demand/cost shifter of international trade flows, $\epsilon_{d\rho}$, and the travel times $T_{d\rho}$.

To solve this endogeneity issue, I use an instrumental variable approach. This approach requires a valid instrument $Z_{d\rho}$. The instrumental variable needs to be relevant, $E[Z_{d\rho}T_{d\rho}] \neq 0$, and exogenous, $E[Z_{d\rho}\epsilon_{d\rho}] = 0$. I consider two instrumental variables: the distance between ports and capitals of departments using the road network of Colombia in 1938, and the distance between city-ports and the capitals of departments using the 17th-century colonial roads of the Viceroyalty of New Granada. These instrumental variables are similar to the ones used by Duranton (2015) to analyze the domestic trade between Colombian cities. I discuss the validity of the instrument below. Duranton, Morrow, and Turner (2014), Baum-Snow (2007), and Michaels (2008) also use a similar approach.

The road network of 1938 served specific regional purposes because railroads played a major role in the transportation of goods. Therefore, the transportation policies implemented by the Colombian national government focused on the expansion of the railroad network (Pachon and Ramirez, 2006; Alvear-Sanin, 2008). Also, as Duranton (2015) pointed out, the road infrastructure did not serve international trade purposes. For example, the two most populated Colombian cities (Medellin and Bogota) did not have a road connection to the Atlantic seaports (see Appendix B).

Duranton (2015) describes with detail the characteristics of the colonial road network (*caminos reales*). Some of the *caminos reales* were used by the indigenous tribes that lived in the country before the Spanish colonizers arrived. They mainly consisted of trails and paths used by the Spanish colonizers to travel to the interior of Colombia. To travel along these trails, it was necessary to use mules. Therefore, Duranton (2015) argues that internal trade was very small within colonial towns. Moreover, the first census implemented in Colombia at the beginning of the 19th century (two centuries after the colonial routes were established) indicates there were less than 2.4 million people in the country (DANE, 2019).[15] According to the 2018 census generated by DANE, Colombia had a population of 48.2 million persons. To sum up, the economic conditions that lead to the establishment of the colonial routes were very different, relative to the current economic circumstances that define which city-port a department uses to trade with the rest of the world.

The distance using an old road network is correlated with the travel times using the current road network, given that it is easier and less costly to build new roads using existing old paths or roads, relative to constructing new roads using new land. The exogeneity of my instrumental variables comes from the fact that given the economic conditions that explain the structure of the old road networks, it is highly likely that the current demand/cost shifters of the trade flows for a pair department city-port, (d, ρ), are uncorrelated with distance using old road networks, given that these network were built when the structure of the Colombian economy was different. In the 17th century, domestic trade in the country was relatively small. During 1938, Colombia was mainly an agricultural economy.

Given that for some department-port pairs, there is not a connection in the old road networks, I created two categorical indices based on the estimated road distances between locations using Dijkstra's algorithm, one for each road network. These indices include a category for those department-port pairs unconnected in the historical road networks. Table 2 reports the results of my estimation, combining both instrumental variables. There is no evidence of weak instrumental variables, given the value of the F-statistic of the first stage (Stock and Yogo, 2005). Moreover, the 2SLS estimates are more precise, compared to the OLS estimates.

[15] The first census of Colombia was implemented in 1822, and included the nations of Venezuela, Panama, Colombia and Ecuador, which were part of the former Republic of Colombia.

Table 1.2 Empirical results of the gravity equation

Method	(1) OLS	(2) 2SLS	(3) OLS	(4) 2SLS	(5) OLS	(6) 2SLS
Sector	agriculture		mining		manufacturing	
$-\mu_{t,k}(time_{dp})$	-0.3282	-0.5274	-0.2293	-0.5199	-0.2880	-0.6182
	(0.3410)	(0.0576)	(0.4100)	(0.0642)	(0.3190)	(0.0592)
F-statistic (1st stage)	-	13.94	-	13.94	-	13.94
N	1,026	1,026	1,026	1,026	1,026	1,026
R-squared	0.5430	0.5230	0.4600	0.4180	0.6910	0.6531

Notes: The categorical variables that I use as instrumental variables have a value of 1 if the department and the port are in the same city; a value of 2 if the distance between the locations is 1-300 kilometers for the 1938 road network, and 1-330 km for the colonial road network; a value of 3 if the distance between locations is 300-700 kilometers using the 1938 road system and 330-830 kilometers using the colonial path system; a value of 4 if the distance is larger than 700 km using 1938 roads, or the distance is longer than 830 kilometers using the 17th century roads; and a value of 5 for those locations unconnected using the old road network.

Given that there is a negative sign multiplying the parameter $\mu_{t,k}$ according to my structural model, then the value of this parameter is positive. Table 2 shows that the magnitude of the OLS estimate is smaller in absolute value, compared to the magnitude of the 2SLS estimate. This implies that any unbservable governments policies that are affecting both exports between department-port pairs and their travel times, are being targeted at regions with large travel times with respect to city-ports (or equivalently, poor infrastructure). This is consistent with the evidence provided by Pachon and Ramirez (2006) and Alvear-Sanin (2008) regarding infrastructure policies in Colombia.

1.4.5. Estimation of the parameter of the dispersion of productivity of shipping routes

To obtain estimates of the parameters $\theta_k \forall k \in \{a, m, i, z\}$, I use estimates of β_{time} from Allen and Arkolakis (2019). The authors consider the function $\tau_{nj} = exp(\beta_{time} T_{nj})$ in their estimation procedure, where T_{nj} is the travel time between locations n and j. They report $\beta_{time} = 0.08$ for a trade elasticity $\sigma = 9$. If I use the elasticity of substitution $\sigma = 6$, then $\beta_{time} = 0.13$. I report my estimates of the parameter θ_k in table 3.

A potential concern is that the estimate of β_{time} comes the context of the American network road system. The empirical evidence of Atkin and Donaldson (2015) shows that the relationship between intra-national trade costs and distance/travel times is very different in developing countries (Ethiopia and Nigeria) relative to the United States.

Although this may represent a concern, there is a caveat. First, data from the World Bank suggests that for the year 2012, Colombia's quality of infrastructure for trade and logistics was much higher compared to the African countries analyzed by Atkin and Donaldson (2015)[16]. This suggests that, even though the values for the parameter β_{time} may not be the same for the United States and Colombia, their differences must be much smaller than the reported by Atkin and Donaldson (2015) between the two African countries and the United States.

As a robustness check, I run my counterfactuals with other values of β_{time}. Specifically, I consider that the parameter can be 10% and 20% higher than the one from Allen and Arkolakis (2019) as it is shown in table 3. For the purpose clarity, Appendix E contains graphs on how the values of the parameters θ_a, θ_m and θ_i vary if I also consider the confidence intervals of my estimates of $\mu_{t,a}$, $\mu_{t,m}$ and $\mu_{t,i}$.

To interpret the magnitudes of the estimate of the parameter θ_k, it is necessary to recall that it is the shape parameter of the Frechet distribution. Economically, it represents the dispersion of the productivities of the international shipping routes (or equivalently, the dispersion of productivities of the city-ports). A high value for θ_k implies low heterogeneity in the productivity of the city-ports to export a good from sector k. A low value, represents high heterogeneity in the productivities of the city-ports. Given the values that I report in Table 3, this implies that for all three sectors, the city-ports show high levels of heterogeneity in their productivities. Intuitively, this implies that firms within a department tend to choose different city-ports to export and import goods from the rest of the world.

Notes: To obtain the values of θ_k, I use the estimates of $\hat{\mu}_{t,k}$ shown in Table 2 for every sector $k \in \{a, m, i, z\}$ and the value of β_{time} from Allen and Arkolakis (2019). Then, I adjust the value of β_{time} upwards.

[16] In 2012, Colombia ranked 64th in the Logistics Performance Index of the World Bank (there are 168 positions). Ethiopia and Nigeria's positions were 141 and 118, respectively. The United States ranked 4th.

Table 1.3 Values for θ_k for different values of β_{time}

Parameter	$\theta_{agriculture}$	θ_{mining}	$\theta_{manufacturing}$
Values when $\beta_{time} = 0.13$ (Allen and Arkolakis, 2019)	4.06	4.00	4.76
Values when $\beta_{time} = 0.143$ (10% higher than baseline)	3.69	3.64	4.32
Values when $\beta_{time} = 0.156$ (20% higher than baseline)	3.38	3.33	3.96

1.5. The impact of *Ruta del Sol* on comparative advantage

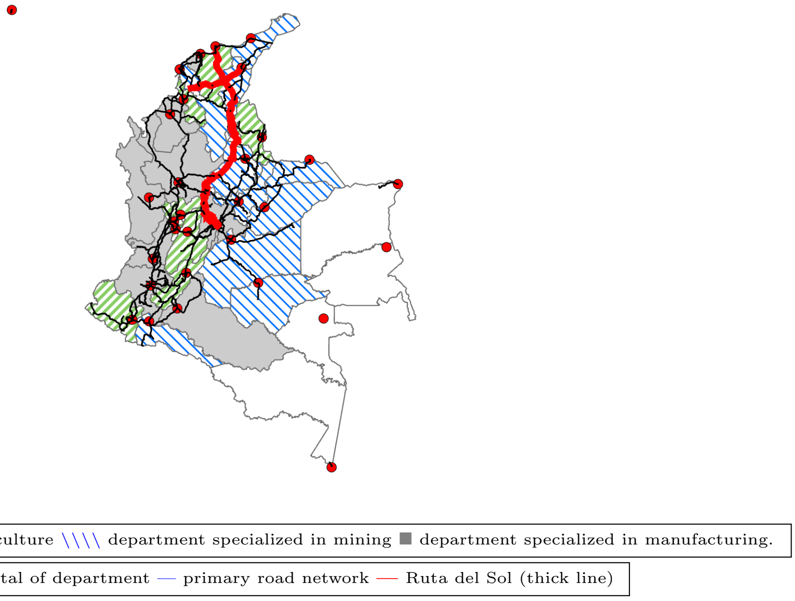
1.5.1. Expected impacts of the road project

I evaluate the effects of the construction of the infrastructure road project *Ruta del Sol*. The project consists of the construction, renovation, and expansion of lanes for 1,071 kilometers of the primary road system. The objective of the highway is to improve the connectivity between the center of the country and the Atlantic Ocean seaports. There was an unsuccessful attempt to start construction in 1997. A decade later, the Colombian government made a second attempt to start the project in 2009.

The project consists of three segments. The bidding process occurred in 2009, and contracts were negotiated and signed the following year (INCO, 2010a; INCO 2010b and INCO 2010c). The beginning of the construction for different segments started in the period 2010-2011. The project has faced multiple delays in its completion, although many sub-segments were inaugurated during the period 2014-2019 as the local media reported (El Espectador, 2019; La Republica, 2014, Semana 2019).

To measure the effects of the infrastructure project on travel times, I create a road network that includes improvements in the segments that already exist and those segments not built yet. I consider that after the completion of the project, the speed of the roads improves from 80 km/hour (approximately 50 miles/hour) to 100 km/hour (approximately 60 miles/hour). I chose a small change in speed derived from the completion of the project for the existing road segments. One of the main objectives of the project is to guarantee the existence of two lanes along the highway. This improvement particularly benefits trucks, by increasing physical maneuverability, particularly in the areas where

Figure 1.7 Location of the project "Ruta del Sol"

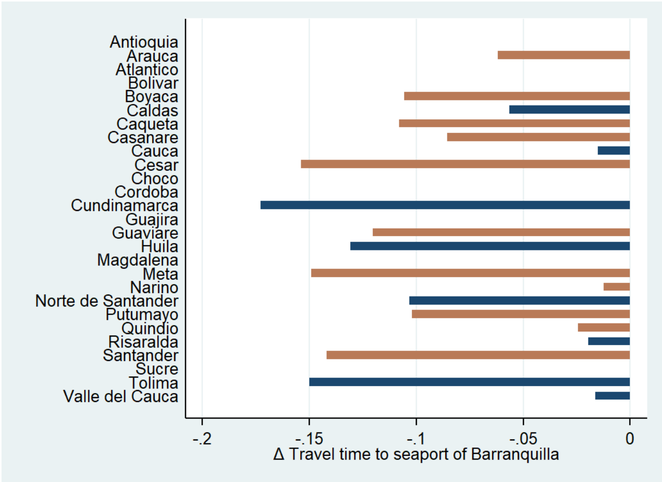


Notes: The colors/figures that fill the area of every department show the sector with highest value of the Balassa index.

the highways cross hilly regions. Such improvement has a direct impact on the speed of vehicles.

A priori, the effects of *Ruta del Sol* on the comparative advantage of Colombia are unclear. Figure 1.7 shows that the road crosses regions that specialize in different sectors. The project improves the connectivity between the department of Cundinamarca, which specializes in manufacturing, and the Atlantic seaports. But also reduces the travel times between departments that sepcialize in mining and the same seaports. Moreover, graphs in the Appendix D show that the international trade costs τ_{dRoW} fall for several departments and all tradable sectors, according to the predictions of my framework.

Figure 1.8 Reduction in travel distance between department and seaport of Barranquilla (%)



● department specialized in mining ● department that does not specialize mining

Notes: The change in travel times is measured as a fraction of the original travel times, without Ruta del Sol. I calculate the specialization of every department with the Balassa Index. I estimate the travel times between the capital of each department (shown in the vertical axis) and the seaport of Barranquilla for a baseline scenario and a new scenario. For the baseline, I assume that the existing segments that already exist (but will be improved) have a speed of 80 km/h. For the scenario in which the project is completed, these existing segments will have a speed of 100 km/h. In addition, I also include the planned new segments. Thus, the new scenario, in which the road project is completed, includes both the new and the improved road segments of *Ruta del Sol*.

1.5.2. Relevance of the parameter that defines relationship between trade costs and travel times

A potential concern regarding the evaluation of the effects of new infrastructure on sectoral exports is the choice of the value of β_{time}, which comes from the American road system. The value of the parameter affects the results given that it determines how changes in travel times in the Colombian primary road system lead to changes in domestic and international trade costs.

Although I do not have the true value of β_{time} for Colombia, there are reasons to believe the value of this parameter is higher in Colombia compared to the United States. This idea is supported by the empirical evidence of Atkin and Donaldson (2015), which suggests that travel times have a larger effect on trade costs in developing nations, relative to the United States' context. As a robustness check, I report the results of simulations using different values of β_{time}. These results are in Appendix E.

Appendix J shows the estimates of the effect of the completion of the highway *Ruta del Sol* in the trade costs $\tau_{dRoW,k}$ for different values of β_{time}. These graphs illustrate that using the value of β_{time} from Allen and Arkolakis (2019) leads to conservative estimates of the change in trade costs caused by the completion of *Ruta del Sol*.

1.5.3. Results of the main simulations

I report the effects of my simulations on the share of agricultural, mining, and manufacturing exports in table 4. As I mentioned before, for small open economies and aggregated sectors, the share of exports for a specific sector is a good proxy to measure shifts in the comparative advantage of a country in a specific sector (this share is the numerator of the Balassa index of RCA). As French (2017) documents, RCA indices are useful to measure patterns of comparative advantage.

As a robustness check, I implement my simulations under different values for the share of expenditures on own goods for the case of domestic trade of Colombian departments, μ_{dd}.[17]

Table 1.4 Results of the simulation under different parameters

	Counterfactual	$\bar{\mu}_{dd}$	$\frac{X_{agriculture,Col}}{X_{total,Col}}$	$\frac{X_{mining,Col}}{X_{total,Col}}$	$\frac{X_{manuf.,Col}}{X_{total,Col}}$	**Change in the share of manuf. exports (p.p.)**
I	No new road	0.3	7.70 %	54.19%	38.10%	
	Completion of Ruta del Sol	0.3	7.53 %	50.15 %	42.32 %	**+4.22**
II	No new road	0.6	7.39%	55.76 %	36.85 %	
	Completion of Ruta del Sol	0.6	7.43 %	51.37 %	41.21 %	**+4.35**

Note: μ_{dd} is the share of expenditures of a department in its own goods for the agricultural and manufacturing sector (only considering the domestic trade flows)

Under different values of the parameters that govern the trade and transportation framework, the results are similar: the infrastructure project *Ruta del Sol* leads to a higher

[17]See section "Taking the model to the data" for more details

share of manufacturing exports, even though it increases the connectivity of several mining departments.

1.5.4. Mechanisms: the role of regions and their specialization

I analyze which departments contribute to the increase in the share of manufacturing exports in my simulations. The improvements in the connectivity of the department of Cundinamarca and the capital district of Bogota could be the main source of this growth. This is because the three main manufacturing regions in Colombia are located in the metropolitan areas of Bogota (capital district and capital of the department of Antioquia), Medellin (capital of the department of Antioquia), and Cali (capital of the department of Valle del Cauca). Nevertheless, among these three manufacturing cities, Bogota is the one that would observe the largest reductions in the travel times to the Atlantic seaports, as figure 1.8 shows.

To examine the regional contributions to the changes in the share of manufacturing exports, I analyze the change in manufacturing exports between two different equilibria using the following expressions

$$\Delta\ Manuf.\ Share\ Exports = \frac{X'_{i,Colombia}}{X'_{Colombia}} - \frac{X_{i,Colombia}}{X_{Colombia}}$$

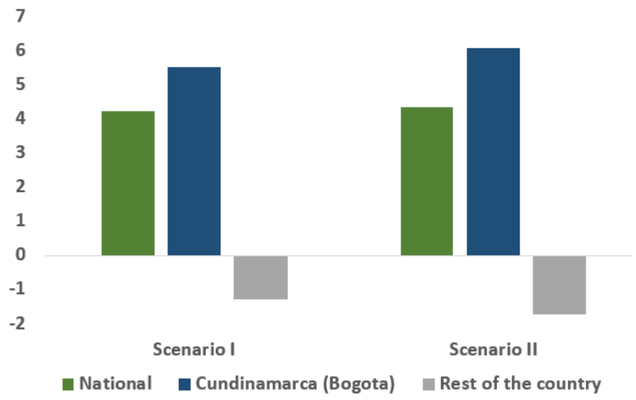
$$\Delta\ Manuf.\ Share\ Exports = \Big[\Big(\frac{X'_{i,b}}{X'_{Colombia}}\Big)+\Big(\frac{\sum_{d\neq b} X'_{d,i}}{X'_{Colombia}}\Big)\Big]-\Big[\Big(\frac{X_{i,b}}{X_{Colombia}}\Big)+\Big(\frac{\sum_{d\neq b} X_{i,d}}{X_{Colombia}}\Big)\Big]$$

$$\Delta\ Manuf.\ Share\ Exports = \Big[\frac{X'_{i,b}}{X'_{Colombia}} - \frac{X_{i,b}}{X_{Colombia}}\Big] + \Big[\frac{\sum_{d\neq b} X'_{i,d}}{X'_{Colombia}} - \frac{\sum_{d\neq b} X_{i,d}}{X_{Colombia}}\Big] \quad (1.30)$$

where $X_{Colombia}$ and $X'_{Colombia}$ are total exports of Colombia under the old and new equilibrium, respectively; $X_{i,d}$ and $X'_{i,d}$ are the manufacturing exports of the department d for the case of the old and new equilibrium, respectively.

Using (30), the change in share of manufacturing exports is the contribution of any department b (first term parenthesis) plus the contribution of the rest of the departments (second term in parenthesis). Hence, I decompose the growth in the share of manufacturing exports for different scenarios. I display the results of the decomposition in figure 1.9. These results show that the increase in manufacturing exports of Cundinamarca is the main driver of the change of national comparative advantage towards manufacturing.

Figure 1.9 Decomposition of growth in share of manufacturing exports (%)



Note: μ_{dd} is the share of expenditures of a department in its own goods for the agricultural and manufacturing sector (only considering the domestic trade flows)

1.5.5. Mechanisms: the role of industry linkages and the structure of the road system

To understand the forces driving the shift of the comparative advantage of Colombia towards manufacturing, I analyze the increase in the share of manufacturing exports under different counterfactual scenarios that consider separately the road network effects of *Ruta del Sol*, with and without input-output linkages.

In the first alternative counterfactual (scenario A), I close the input-output linkages but I allow for the impact of the road infrastructure project on both domestic and international trade costs (see equation 5). This implies that firms producing intermediate goods exclusively use labor as input. The second alternative counterfactual (scenario B) allows for the existence of input-output linkages, but only takes into account the effects of *Ruta del Sol* on the domestic trade costs. Lastly, I run a third alternative counterfactual simulation (scenario C), in which I consider industry linkages, but I assume the road project only affects international trade costs, and does not change domestic trade costs. I report the results of these alternative counterfactual experiments in columns 2 and 3 of table 5.

Table 1.5 Results of alternative simulations

	Scenario	**Increase in the share of manufacturing exports**	
		($\bar{\mu}_{dd} = 0.3$)	($\bar{\mu}_{dd} = 0.6$)
Main	All the effects of *Ruta del Sol*	+4.2	+ 4.4
A	Impacts of *Ruta del Sol* without considering input-output linkages	+1.2	+1.8
B	*Ruta del Sol* only affects domestic trade costs	+0.7	+0.7
C	*Ruta del Sol* only affects international trade costs	+3.6	+2.0

Note: When I use the Balassa index for Colombia, the share of exports of a sector is a good proxy of its comparative advantage, since the denominator of the index is given for a small open economy and if the sectors are not very disaggregated. μ_{dd} is the share of expenditures of a department in its own goods for the agricultural and manufacturing sector (only considering the domestic trade flows).

The alternative counterfactual simulations provide two interesting insights about the forces driving my results. The first insight is that improvements in infrastructure lead to better access to intermediate inputs. This specially benefits manufacturing exports. To see this, the results of the scenario B are informative. In this alternative simulation, I consider that *Ruta del Sol* only improves access to domestic inputs. As a result, the national share of manufacturing exports increases by 0.7 percentage points. Hence, the improvement in the access of domestic inputs alone helps to increase the share manufacturing exports.

The second insight is that the existence of industry linkages propagate the positive effects generated by the road project. The presence of such linkages benefit the manufacturing sector the most. To see this, notice that in scenario A, in which input-output linkages are not considered, the reductions in trade costs lead to an increase in the specialization of Colombia in manufacturing goods, but this growth is one third of the increase from the main counterfactual (scenario A vs. main scenario).

The alternative counterfactuals show the relevance of industry linkages when we measure the impact of road projects, using international trade models. These linkages are not usually considered in existing studies regarding the general equilibrium effects of infrastructure improvements. Failure to consider these linkages will result in the estimation of smaller effects of lower trade costs on the trade flows of specific sectors.

1.6. Conclusion

The main conclusion of this paper is that domestic trade costs are determinants of comparative advantage. This idea is especially relevant for those countries with low quality of infrastructure. Quality of roads influence the spatial distribution of trade costs, thus influencing the availability of factor endowments and inputs for the production of goods and services across regions within an economy. Hence, to have a more comprehensive view of the comparative advantage of a country, it is necessary to consider the structure of its road system.

This idea also has policy implications. Infrastructure projects can be a tool for those policymakers whose objective is to shift the comparative advantage of a country in a particular direction. Given that reductions in trade costs lead to welfare gains through multiple channels, as recent economic literature predicts, the construction of roads seems to be a feasible policy alternative to change the national comparative advantage.

Specifically, in the context of Colombia, one of the most important infrastructure policy projects *Ruta del Sol*, has the potential to change the comparative advantage of the country, by weakening the comparative advantage of the country in mining goods, while strengthening the comparative advantage of the nation in the manufacturing sector. My results indicate that the share of manufacturing exports would grow 4 percentage points in the long run due to the completion of the project. The importance of this magnitude is supported by the fact that in the past three decades, the share of exports for two mining products (coal and oil) has observed a substantial increase, from 30% in 1992, to 58% in 2018. Moreover, the reduction of the concentration of Colombian exports in mining goods is aligned with the objectives of public officials.

The change in the comparative advantage of Colombia caused by *Ruta del Sol* is driven by two forces. First, the road project increases access to global markets of the department of Cundinamarca, which specializes in manufacturing goods. Second, the improvement in access to inputs benefits the manufacturing firms the most, given the structure of input-output linkages of the country.

Lastly, my results highlight the relevance of input-output linkages when considering how infrastructure shapes comparative advantage. I show that when industry linkages are not considered, the increase in the share of manufacturing exports is one third of the growth observed for simulations that consider industry linkages. This result is specially relevant for previous work regarding the economic effects of infrastructure projects, given that little attention has been paid to the relationship between infrastructure and input-output linkages.

1.7. Bibliography

1. Acemoglu, D. and F. Zilibotti (1997). "Was Prometheus Unbounded by Chance? Risk, Diversification and Growth". *Journal of Political Economy*, Vol. 105, No. 4, 709-751.

2. Alder, S. (2019). "Chinese Roads in India: The Effect of Transport Infrastructure on Economic Development", mimeo, University of North Carolina, Chapel Hill.

3. ANIF (2014). "Costos de Transporte, Multimodalismo y la Competitividad de Colombia". Centro de Estudios Economicos de la Asociación Nacional de Instituciones Financieras. Reporte elaborado para la Cámara Colombiana de la Infraestructura. Bogota, D.C., Colombia.

4. Allcott, H. and D. Keniston (2018). "Dutch Disease or Agglomeration? The Local Economic Effects of Natural Resource Booms in Modern America". *Review of Economic Studies*. Vol. 85, No. 2, pp. 596-731

5. Allen, T. and D. Atkin (2016). "Volatility and the Gains from Trade". NBER Working Paper No. 22276.

6. Allen, T. and C. Arkolakis (2019). "The Welfare Effects of Transportation Infrastructure Improvements". NBER Working Paper No. 25487

7. Alvear-Sanin, J. (2008). "Historia del Transporte y la Infraestructura en Colombia, 1492-2007", Ministerio de Transporte, Bogota D.C., 1a Edicion: Diciembre de 2008.

8. Arkolakis, C., S.K. Lee and M. Peters (2018). "European Immigrants and the United States' Rise to the Technological Frontier in the 19th Century". 2019 Meeting Papers 1420, Society for Economic Dynamics.

9. Atkin, D. and D. Donaldson (2015). "Who's Getting Globalized? The Size and Implications of Intra-National Trade Costs", NBER Working Paper 21439.

10. Bahar D. and Rapoport H. (2018). "Migration, Knowledge Diffusion and the Comparative Advantage of Nations", *The Economic Journal*

11. Balassa, B. (1965). "Trade liberalization and 'revealed' comparative advantage". *The Manchester School.* Vol. 33, Issue 2, pp. 99-123.

12. Bartelme, D., A. Costinot, D. Donaldson and A. Rodriguez-Clare (2019). "The Textbook Case for Industrial Policy: Theory Meets Data", mimeo University of Michigan.

13. Baum-Snow, N. (2007). "Did Highways Cause Suburbanization?", The Quarterly Journal of Economics, Vol. 122, Issue 2, pp. 775-805.

14. CAF (1993). "Proyectos Viales de Integracion Andina". Corporacion Andina de Fomento, Santafe de Bogota, Enero 1993.

15. Caliendo, L. and F. Parro (2015). "Estimates of the Trade and Welfare Effects of Nafta". *The Review of Economic Studies*, Vol. 82, No. 1, pp. 1-44.

16. Caselli, F., M. Koren, M. Lisicky, and S. Tenreyro (2019). "Diversification through Trade. The Quarterly Journal of Economics", *The Quarterly Journal of Economics* forthcoming.

17. Chandra, V., J. Boccardo and I. Osorio (2007). "Export Diversification and Competitiveness in Developing Countries", mimeo, World Bank.

18. Coatsworth, J. (1979). "Indispensable Railroads in a Backward Economy: The Case of Mexico". *The Journal of Economic History*, Vol. 39, No. 4, pp. 939-960.

19. Cosar, K. and P. Fajgelbaum (2016). "Internal Geography, International Trade and Regional Specialization". *American Economic Journal: Microeconomics*, 8(1): 24-56.

20. DANE (2019). "DANE en el Bicentenario: Historia de la división político-administrativa y los censos en Colombia". Available online https://www.dane.gov.co/index.php/estadisticas-

por-tema/demografia-y-poblacion/censo-nacional-de-poblacion-y-vivenda-2018/cuantos-somos

21. Deardoff, A. (2014). "Local comparative advantage: Trade costs and the pattern of trade". International Journal of Economic Theory, Vol. 10, Issue 1, pp. 9-35.

22. Dekle, R., J. Eaton and S. Kortum (2008). Global Rebalancing with Gravity: Measuring the Burden of Adjustment. *IMF Staff Papers*, Vol. 55, No. 3, pp. 511-540.

23. DiGiovanni, J. and A. Levchenko (2012). NBER International Seminar on Macroeconomics, University of Chicago Press, vol. 8(1), pages 97-151.

24. DNP (2014). "Logística: Factor Clave para el Comercio Exterior". Presentation by the Director of Infrastructure and Sustainable Energy Dimitri Zaninovich Victoria. Available online https://www.dnp.gov.co/DNPN/direcciones/dirección-de-infraestructura-y-energia-sostenible.

25. Donaldson, D. (2018). "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure". *American Economic Review*, Vol. 108 (4-5), pp. 899-934.

26. Donaldson, D. and R. Hornbeck (2016). "Railroads and American Economic Growth: A 'Market Access' Approach". *The Quarterly Journal of Economics*, Vol. 131, Issue 2, pp. 799-858.

27. Ducruet, C., R. Juhasz, D. Nagy and C. Steinwender (2019). "All aboard: The Aggregate Effects of Port Development", mimeo, MIT Sloan.

28. Duranton, G., P. Morrow, and M.A. Turner (2014). "Roads and Trade: Evidence from the US". *The Review of Economic Studies.* 81 (2), 681–724.

29. Duranton, G. (2015). "Roads and Trade in Colombia". *Economics of Transportation*, Vol. 4, pp. 16-36.

30. Eaton, J. and S. Kortum (2002). "Technology, Geography and Trade", *Econometrica*, Vol. 70, Issue 5, September 2002, pp. 1741-1779.

31. El Espectador (2014). "Primer tramo de la Ruta del Sol, en veremos". October 14th, 2014. Available online. https://www.elespectador.com/noticias/bogota/primer-tramo-de-ruta-del-sol-veremos-articulo-522631

32. El Tiempo (2007). "Si de Procuraduria al caso Commsa", Redaccion El Tiempo, August 15th, 2007. Available online in https://www.eltiempo.com/archivo/documento/MAM-2618824

33. Faber, B. (2014). "Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System", *Review of Economic Studies*, Vol. 81(3), pp. 1046-70.

34. Faber, B. and C. Gaubert (2019). "Tourism and Economic Development: Evidence from Mexico's Coastline". *American Economic Review*, Vol. 109, No. 6, June 2019, pp. 2245-93.

35. FAO (2002). "Dependence on Single Agricultural Commodity Exports in Developing Countries: Magnitude and Trends", mimeo, discussion paper presented at the FAO Round Table on Special and Differential Treatment, WTO Negotiations on Agriculture, Geneva, February 1st 2002.

36. Fajgelbaum, P. and S. Redding (2018). "Trade, Structural Transformation and Development: Evidence from Argentina 1869-1914". NBER Working Paper 20217.

37. FEDESARROLLO (2015). Plan Maestro de Transporte Intermodal (PMTI) 2015-2035: Infraestructura Para el Comercio Exterior, el Desarrollo Regional y la Integración del Territorio. Informe Presentado a Financiera de Desarrollo Nacional.

38. French, S. (2017). "Revealed comparative advantage: What is it good for?". *Journal of International Economics*, Vol. 106, pp. 83-103.

39. Fogel, R. (1962). "A Quantitative Approach to the Study of Railroads in American Economic Growth: A Report of Some Preliminary Findings". *The Journal of Economic History*, Vol. 22, No. 2, pp. 163-197.

40. Hanson, G.H., N. Lind and M.A. Muendler (2015). "The Dynamics of Comparative Advantage", NBER Working Paper 21753.

41. Holl, A. (2016). "Highways and Productivity in Manufacturing Firms". *Journal of Urban Economics*, Vol. 93, pp. 131-151.

42. IADB (2014). "Rethinking Productive Development: Sound Policies and Institutions for Economic Transformation". Edited by Gustavo Crespi, Eduardo Fernández-Arias, and Ernesto Stein.

43. IGAC (2002). "The Atlas of Colombia 2002". Instituto Geográfico Agustín Codazzi. Santa Fé, Bógota, 5a edición revisada y aumentada.

44. INCO(2010a). "Acta de Inicio del Contrato de Concesion No 002 del 14 de Enero de 2010, Proyecto Ruta del Sol - Sector 1: Tobiagrande/Villeta - El Korán". Legal document of the Instituto Nacional de Concesiones, República de Colombia.

45. INCO(2010b). "Acta de Inicio del Contrato de Concesion No 001 del 14 de Enero de 2010, Proyecto Ruta del Sol - Sector 2: Puerto Salgar - San Roque". Legal document of the Instituto Nacional de Concesiones, República de Colombia.

46. INCO(2010c). "Acta de Inicio del Contrato de Concesion No 001 del 4 de Agosto de 2010". Legal document of the Instituto Nacional de Concesiones, República de Colombia.

47. Kremer, M. and S. Jayachandran (2006). "Odious debt", *The American Economic Review*, Vol. 96, No. 1, March 2006, pp. 82-92.

48. Koren, M. and S. Tenreyro, (2007). "Volatility and Development". *The Quarterly Journal of Economics*, Vol. 122, No. 1, 243-287.

49. Kruger, A. B. Tuncer (1982). "An Empirical Test of the Infant Industry Argument". *The American Economic Review*, Vol. 72, No. 5, December 1982, pp. 1142-1152.

50. Krugman, P. (1987). "The narrowing moving band, the Dutch disease, and the competitive consequences of Mrs. Thatcher: Notes on trade in the presence of

dynamic scale economies", *The Journal of Development Economies*, Vol. 27, Issues 1-2, October 1987, pp. 41-55.

51. La Republica (2014). "Mintransporte inauguró obra en el Sector 2 de la Ruta del Sol". May 3rd, 2014. Available online https://www.larepublica.co/infraestructura/mintransporte-inauguro-obra-en-el-sector-2-de-la-ruta-del-sol-2117241

52. Levchenko, A. (2005). "Financial Liberalization and Consumption Volatility in Developing Countries". *IMF Staff Papers*, 52:2, 237-259.

53. Levchenko, A. (2007). Institutional Quality and International Trade. *The Review of Economic Studies*, Vol. 74, No. 3, pp. 791-819.

54. Levchenko, A. and J. Zhang (2016). "The Evolution of Comparative Advantage: Measurement and Welfare Implications". *Journal of Monetary Economics*, Vol. 78, pp. 96-111.

55. Matsuyama, K. (1992). "Agricultural Productivity, Comparative Advantage and Economic Growth", *Journal of Economic Theory*, 58, pp. 317-334.

56. Michaels, G. (2008). The Effect of Trade on the Demand for Skill - Evidence from the Interstate Highway System. *The Review of Economics and Statistics*, Vol. 90, No. 4, pp. 683-701.

57. Ministerio de Transporte (2015). "Transporte en Cifras. Estadisticas 2015". Oficina Asesora de Planeacion, Grupo de Planificacion Sectorial.

58. MIT Media Lab (2019). "The Observatory of Economic Complexity". Available online https://atlas.media.mit.edu

59. Morales, N. (2019). "High-Skill, Multinational Companies and the Location of Economic Activity", mimeo, University of Michigan.

60. Morando, B. (2019). The Reversal of Comparative Advantage? Transportation Costs and Cropping Patterns in Uganda, mimeo, Trinity College Dublin.

61. Oxford Economics (2017). Global Infrastructure Outlook: Infrastructure Investment Needs 50 Countries, 7 Sectors to 2040 Available online https://outlook.gihub.org/

62. Pachon, A. and M.T. Ramirez (2006). "La infraestructura de transporte en Colombia durante el siglo XX". Bogota: Banco de la Republica. Ediciones Fondo de Cultura Economica.

63. Pack, H. and Saggi K. (2006). "Is there a case for industrial policy? A critical survey (English)". *The World Bank research observer*, Vol. 21, No. 2, 267-297.

64. Pacione, M. (2005). "City profile: Dubai". *Cities*, Vol. 22, Issue 3, June 2005, pp. 255-265.

65. Pellegrina, H. and S. Sotelo (2019). "Migration, Specialization and Trade: Evidence from the Brazilian March to the West", mimeo, University of Michigan.

66. Perez-Cervantes, P. (2014). "Railroads and Economic Growth: A Trade Policy Approach", Banco de Mexico Working Paper 2014-14

67. Semana (2006). "Arreglo de Commsa: ¿no era mejor un buen pleito?", September 12th, 2006. Available online https://www.semana.com/on-line/articulo/arreglo-commsa-no-mejor-buen-pleito/80948-3

68. Semana (2006). "El calvario de Commsa", November 18th, 2006. Available online https://www.semana.com/nacion/articulo/el-calvario-commsa/82096-3

69. Semana (2019). "La tragedia de la ruta del Sol", Reportaje Especial. Available online https://especiales.semana.com/ruta-del-sol/

70. Sotelo, S. (2019a). "Domestic Trade Frictions and Agriculture". *Journal of Political Economy*, forthcoming

71. Timmer, M. P., Dietzenbacher, E., Los, B., Stehrer, R. and de Vries, G. J. (2015), "An Illustrated User Guide to the World Input–Output Database: the Case of Global Automotive Production", *Review of International Economics*, Vol. 23, pp. 575–605

72. U.S. Energy Information Administration (2019). "Country Analysis Executive Summary: Colombia". Last updated: January 7th, 2019.

73. Van der Ploeg, F. (2011). "Natural Resources: Curse or Blessing?". *Journal of Economic Literature*, Vol. 49, pp 366-420.

74. Vieira-Posada, E. (2008). "La Formacion de Espacios Regionales en la Integracion de America Latina", 1a Edicion, Pontificia Universidad Javeriana: Convenio Andres Bello.

75. Xu, Hangtian (2016). "Domestic railroad infrastructure and exports: Evidence from the Silk Route". *China Economic Review*, Vol. 41, 129-147.

76. Xu, Mingzhi (2018). "Riding on the New Silk Road: Quantifying the Welfare Gains from High-Speed Railways", mimeo, University of California, Davis.

Chapter 2. The Geography of Commodity Booms

2.1. Introduction

Commodity booms can influence the allocation of resources across and within nations. A large economics literature has focused in how commodity booms affect national outcomes, such as the exchange rate, economic growth and the national manufacturing exports. However, the allocation of one sector's resources across regions has remained understudied. Such aspect is highly relevant since commodity booms have been associated with the creation of new cities (e.g. San Francisco, Seattle, Manaus) or the industrialization of some regions (e.g. Detroit in the United States). Hence, understanding the spatial effects of commodity booms is important to comprehend the structural transformation of nations.

This paper answers two research questions regarding the economic geography of commodity booms: What are the short-run impacts of an oil boom in the regional manufacturing outcomes? Which general equilibrium factors are relevant in the short-run to explain these spatial effects? Historically, some regions have seen growth in the manufacturing sector after experiencing a commodity boom, while others have experienced zero or negative impacts on local manufacturing outcomes. My empirical findings provide evidence that an oil boom leads to a small growth in the local manufacturing output of Colombian departments and positive effects in investment and expenditure on intermediate inputs[1]. I show that these results are caused by the propagation of the shocks through industrial linkages and the presence of high domestic transportation costs in Colombia. IO linkages can generate higher manufacturing output in some sectors and the presence of high transportation costs can increase the demand for local manufacturing goods.

The impacts of oil production in the regional manufacturing goods are also influenced by international domestic trade costs, especially for sectors for which global logistics are more complex due to the nature of the goods (e.g. size, durability, resistance of the materials, weight, special requirements for transportation, etcetera). Hence, an oil boom would lead to increases in the local demand of manufacturing goods for some sectors, overriding the negative impacts on manufacturing output caused by increases in wages and the prices of services. Finally, I provide evidence that the increase in industrial output can occur in the short run, a result highly relevant for the extensive *resource curse* literature that also analyzes the short-run impacts of commodity booms. The mechanisms and timing of the effects I focus on are very different from previous literature that exclusively consider the medium and long-run effects of agglomeration of labor, Allscott and Keniston (2017).

My empirical analysis show that industries related to "Food, beverages and tobbaco" and "Chemical industries" benefit from regional oil production in the short-run. The industries that are negatively impacted by an oil boom tend to be those with fewer linkages to the extraction of oil, and for goods that are easier to transport internationally ("Electrical, mechanical and transportation devices", "Other basic manufacturing industries"). More importantly, there is no evidence that the oil booms generate an increase of manufacturing establishments in the short-run. Therefore, there is no evidence that the effects I find are generated from increases in local demand generated by more migration to oil regions.

These empirical findings are relevant for other contexts. First, they provide evidence that oil production can indeed generate reallocation of resources through mechanisms unrelated to the exchange rates. Second, they illustrate that commodity booms can generate

[1]The department is the administrative region of Colombia. It is equivalent to a state or a province.

positive effects in some manufacturing sectors under specific economic conditions. Third, they highlight the relevance of input-output linkages in the structural transformation of regional economies. Fourthly, the empirical findings highlight that short-run spatial impacts of commodity booms have a significant size. Lastly, they provide evidence that domestic transportation costs are key to understand how a commodity boom transforms a nation. A key element of my findings is that I focus on the short-run, which implies the effects I analyze are substantially different than the medium and long-run effects of agglomeration.

The regional analysis of commodity booms is intricate due to three elements: lack of cross sectional variation of the commodity booms across heterogenous regions within a country, data availability, and endogeneity concerns. Recent research regarding commodity booms focuses on countries where the impact occurs in a few regions (e.g. Peru , Brazil). Another challenge is that large commodity booms might occur in countries with homogeneous geography regions (e.g. Middle East countries). For those cases in which commodity booms show cross sectional variation across heterogeneous regions, the data is relatively difficult to access since these commodity booms occurred in the 19th century or the early 20th century (e.g. tea boom in Asian nations, or the rubber booms in Brazil or Malaysia). Finally, endogeneity is another obstacle. Since unobserved local economic characteristics tend to influence the production of the commodity at a regional level, it is necessary to construct an instrument that exogenously shifts production across time and space. Finding such instruments can be a challenging task.

To address these issues, the context of the oil production boom in Colombia is ideal. First, oil production in the country occurs across departments with different levels of industrialization. Therefore, I am able to analyze how an oil boom impacts heterogeneous regions. Second, Colombia's data regarding the manufacturing sector is highly detailed. This allows me to do a detailed analysis of the impacts of the oil boom in the local manufacturing outcomes. Colombian data has been used in other papers (see Eaton at al, 2007; Eslava et al, 2013; and Fieler et al, 2018 for some examples). Third, using an oil boom deals with endogeneity issues given the fact that oil production exclusively depends on engineering issues, as the environmental economics literature has documented with detail (Kellogg 2011; Anderson, Kellogg and Salant, 2018). This allows me to considerate the oil production in a Colombian department as exogenous to the unobserved time-variant local economic conditions

This paper contributes to the literature regarding the impacts of natural resource booms on national output, economic growth and the manufacturing sector. This literature typically focuses on either the dynamic impacts of natural resource booms, a phenomenon known as the *Dutch Disease*, (see Bjornland et al, 2019 for a review of this work; or Corden and Neary, 1982, and Bruno and Sachs, 1982 for the seminal papers on the topic) or the cross-country analysis of how abundance of natural resources impacts on national economic outcomes (Cassidy, 2018; Caselli and Tesei, 2018; Michaels, 2011; Arezki et al, 2016; Bruckner et al, 2012). I differ substantially from this literature by focusing on the spatial effects of commodity booms.

Focusing on the spatial economic effects of commodity booms is relatively new in the economics literature and has been only explored by Allscott and Keniston (2018) and Feyrer, Mansur and Sacerdote (2017), for the case of oil production in the United States (relatively small with respect to the size of the national economy). I differ from these papers in three aspects. I focus on a developing country, I analyze an oil boom that is very large relative to the size of the regional and national economies, and, more importantly, I provide evidence that the spatial effects of commodity booms can occur in the short-run, and not only in the medium and long-run as Allscott and Keniston (2018) document.

One of the main contributions of this paper is to show that input-output linkages are key to understand the propagation of large regional shocks in an economy. Input-output linkages have been mostly used to analyze international trade issues (Caliendo and Parro, 2014; Antras and De Gortari 2020; Yi, 2003; Costinot, Vogel and Wang, 2013; Barrot and Sauvagnat, 2016; Boehm, Flaaen, and Pandalai-Nayar, 2019; Baqaee and Farhi, 2019; Huo et al, 2010), or the propagations of distortions (Balashundaram, 2019; Baqaee, 2018; Carvalho et al., 2016; Grassi, 2017). Finally, previous empirical work on the local effects of commodity booms have not considered the role of IO linkages.

Finally, my work adds to the development economics literature. Specifically, this paper helps to understand the relationship between commodity booms and structural transformations across regions in a nation. Different from previous work I highlight the relevance of the general equilibrium effects mechanisms to understand the consequences of commodity booms. Traditionally, the development economics literature regarding the local impacts of commodity booms focuses in the *Resource Curse*, the existence of negative political and economic local outcomes after a commodity boom (Ishak and Meon, 2020; Litschig, 2012; Monteiro and Ferraz, 2012; Aragon and Rud, 2013; Brollo et al, 2013; Caselli and Michaels, 2013; Dube and Vargas, 2013; Aragon and Rud, 2016; Carreri and Dube, 2017). Nevertheless, this literature omits general equilibrium effects and by doing so, the external validity of these empirical exercises becomes weaker. One of the conclusions of my paper is to show that general equilibrium effects are highly relevant, hence, the direction and size of the impacts of commodity booms might depend on the IO linkages, the domestic trade costs and the sectoral composition of the regions.

The rest of the paper is organized as follows. Section 2 describes the mechanisms of shock propagation of commodity booms. Section 3 describes briefly the context of the Colombian oil boom. Section 4 describes the data. Section 4 describes the empirical model and discusses identification. Section 5 reports the empirical results. Section 6 concludes.

2.2. Spatial effects of commodity booms: an simple model

I consider two mechanisms that help to propagate spatially the economic impacts of an oil boom in the short-run (and in general, of any commodity boom): IO linkages within the country, and domestic trade costs. These factors are the reason why a commodity boom in a single region can generate very different outcomes. I explain the role of these factors below. I abstain from considering migration, since I am focusing on the short-run effects of commodity oil booms.

Increases in local demand of goods by new households. When a commodity boom increases the local economic activity in a region, this attracts new workers and entrepreneurs. The movement of workers is generated by the increase in the local wages caused by a sudden increase in the demand of labor. When the arrival of new workers and entrepreneurs is large enough, new firms will appear due to more entrepreneurs and higher demand of local goods.

Input-output linkages. A large increase in the production of commodities generates two effects. It increases the demand of goods in those sectors that supply inputs to the sector that is going through a boom (e.g. oil sector requires chemical inputs), and it reduces the cost of inputs for downstream industries (e.g. higher local oil production reduces the unit costs in refineries). A typical feature of a commodity boom is that the services sector tends to grow substantially, due to higher demand of services. Therefore, the local manufacturing industries that supply the services sector might grow.

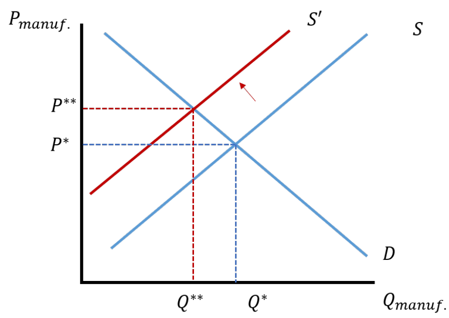
Domestic trade costs. As mentioned previously, commodity booms lead to higher demand of manfacturing goods that are suppliers of the oil sector and the services sector. But in the presence of very low domestic trade costs, the oil region might simply buy such goods from other regions. Hence, under such case there might be a negative impact in the local manufacturing output.

To see the relevance of every one of these factors, the following scenarios illustrate how an oil boom generates different outcomes on the local manufacturing output.

- *Scenario 1. Local manufacturing has no IO linkages with the oil sector and domestic trade costs are high.* In this case an oil boom leads to a large increase in the demand of workers, which pushes wages upwards. This implies higher unit costs in the manufacturing sector (assuming a Cobb-Douglas production function). This results into lower manufacturing output.

- *Scenario 2. Local manufacturing strong IO linkages with the oil sector and domestic trade costs are high.* In this scenario, there are two consequences. Higher local wages due to increases in the demand of workers, and higher demand for local manufacturing goods. Therefore, both the demand and supply of manufacturing

Figure 2.1 Impacts of oil boom in the manufacturing sector (no IO linkages and high domestic trade costs)



goods shift. This can generate higher or lower local manufacturing output (figures 2.2 and 2.3). In some cases, it might generate no changes in output and a simple increase in prices (figure 2.3).

These results depend strongly on the size of the shift in the demand of the local manufacturing goods, the composition of the local industry and the impact of higher wages in the unit costs of the manufacturing sector.

Figure 2.2 Positive impacts of oil boom in the manufacturing sector under scenario 2.

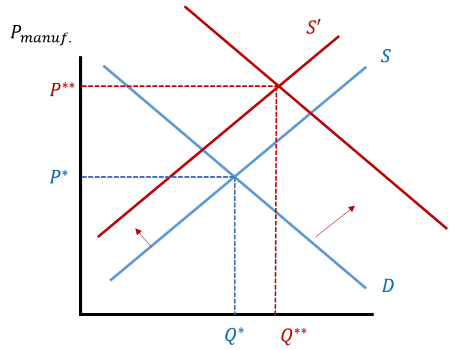
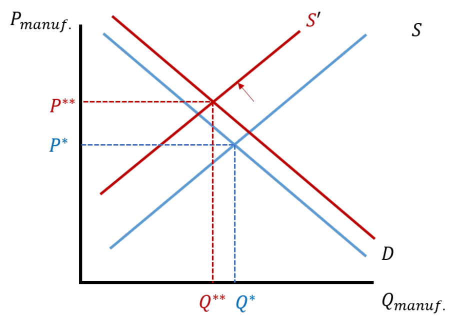
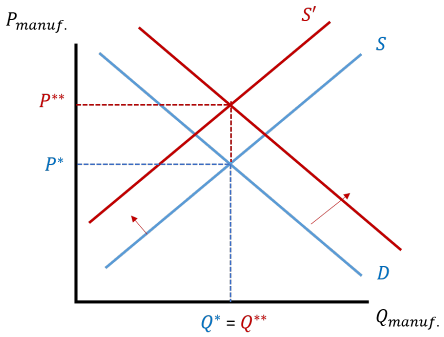


Figure 2.3 Negative impacts of oil boom in the manufacturing sector under scenario 2



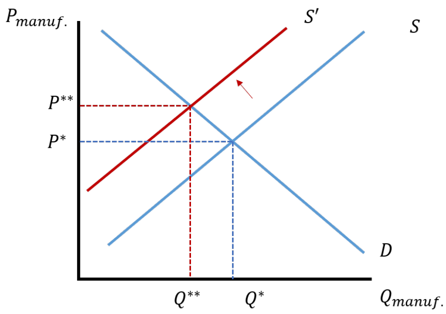
- *Scenario 3. Domestic trade costs are very low.* Given that local wages increase due to a higher demand of workers, the local manufacturing sector cannot compete with manufacturing in neighboring regions. Therefore, independently of the IO linkages,

Figure 2.4 Zero impact of oil boom in the manufacturing output and positive impact on prices, under scenario 2



the region with the oil boom will simply import manufacturing goods from other regions. This leads to lower local manufacturing output.

Figure 2.5 Negative impacts of oil boom in the manufacturing sector under scenario 2



2.3. Empirical facts about the Colombian oil boom

Below, I present four empirical facts regarding the Colombian oil boom. These facts are helpful to understand the national and regional impacts. Moreover they provide evidence that this context is ideal to understand how an oil boom can impact the spatial allocation of resources given that oil production shows both spatial and time variation across Colombian regions. **Fact 1. The oil boom is large relative to the size of the Colombian economy**. Oil production has increased significantly in Colombia since 2004. Since Colombia is an oil exporter, I can use oil exports as a proxy of the value of oil production. Figure 2.4 shows that the sector has experienced a large boom, and then a medium contraction in 2015. Nevertheless, oil production remains high in 2018, relative to 2004 levels (more than 3 times larger).

Figure 2.5 shows that oil has remained as one of the most important exports since

2010 (with the exception of 2016). Therefore, at a national level the oil sector became large enough to generate significant macroeconomics impacts, as Sarmiento and Lopez (2016) document.

Figure 2.6 Value of oil exports in Colombia (billions of current dollars)

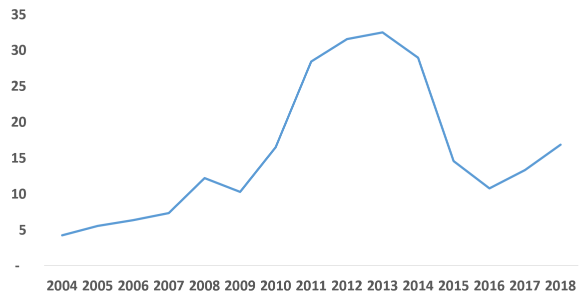
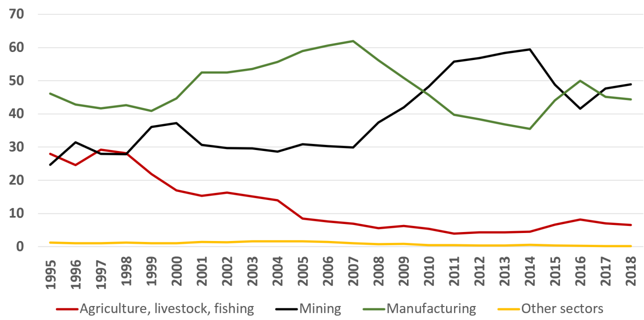
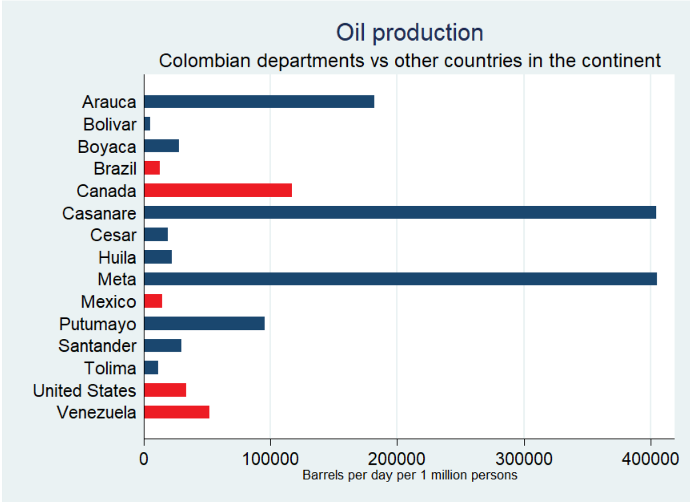


Figure 2.7 Value of oil exports in Colombia (billions of current dollars)



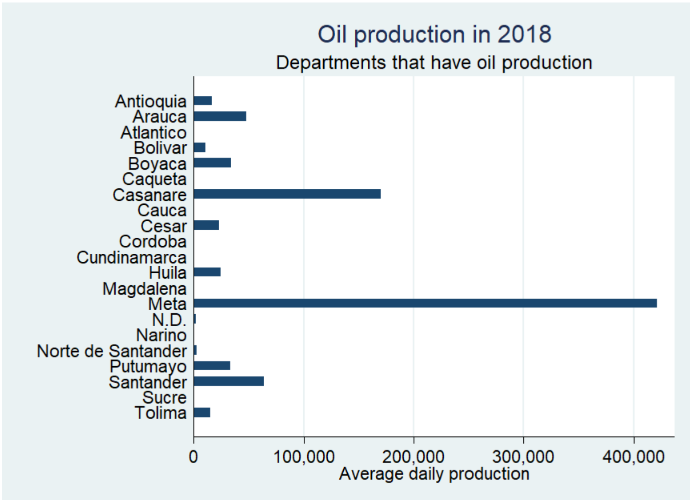
Fact 2. The oil production in many Colombian departments is high for international standards. Using data on oil production per capita, it is noticeable that the local oil production per capita in Colombian departments is as high as the oil production among countries in the American continent such as Brazil, Mexico or Venezuela. In some cases, the oil production of some departments is higher than the the main oil country producers in the Americas.

Figure 2.8 Value of oil exports in Colombia (billions of current dollars)



Fact 3. There is variation in the production of oil across Colombian regions across time and space. One of the main characteristics of this boom is that, although two Colombian departments are responsible for almost half of the oil production in the country (Meta and Casanare concentrate 48% of the national oil production), other departments do have local oil production, as figure 2.6 shows.

Figure 2.9 Value of oil exports in Colombia, 2018 (billions of current dollars)



Moreover, across time it can be noticed that every region follows different dynamics in their local oil production. Figures 2.7 and 2.8 show the evolution of oil production for the ten departments with more oil production across 14 years. It is possible to notice different trends across these departments. These include the upward trends in production in Putumayo, downward trends in Arauca and Huila, and cases where there are both booms and busts in the local oil production as it is the case of Antioquia or Casanare.

Figure 2.10 Average daily oil production among Colombian departments (medium producers, more than 50,000 barrels per day)

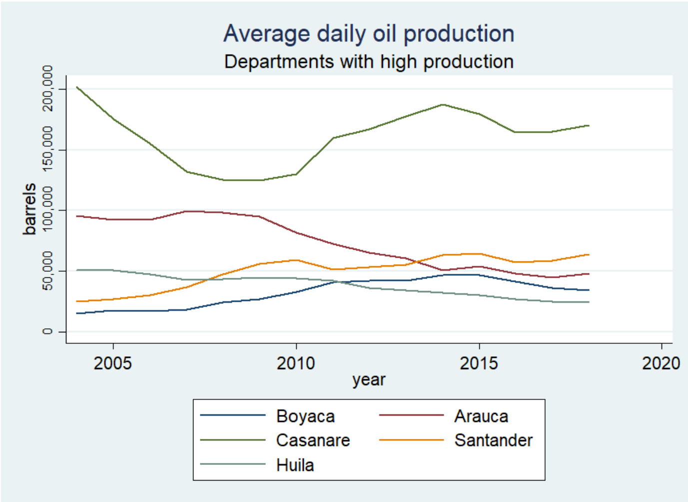
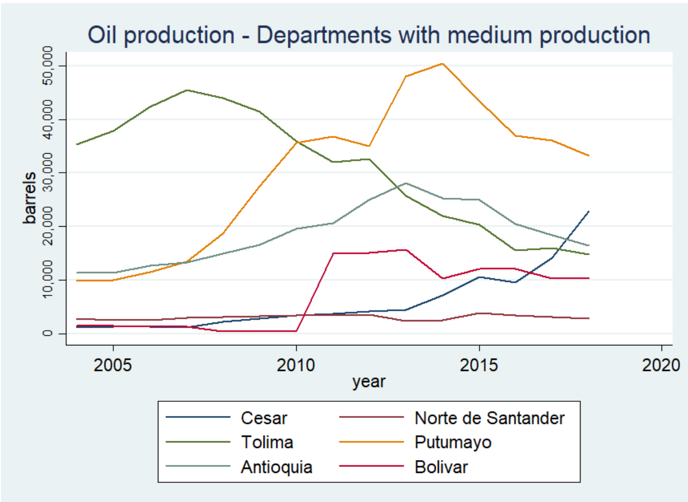


Figure 2.11 Average daily oil production among Colombian departments (medium producers, less than 50,000 barrels per day)



2.4. Data

To analyze the impacts of the oil production, I use two separate datasets. The first dataset is produced by the *Ministerio de Minas y Energía* (Ministry of Mines and Energy) and provides data on the daily average of oil production, per year at a department level, from 2004 to 2018. The data of the Ministry includes the fiscalized oil production and does not include illegal extraction of oil.

The second dataset is the Annual Manufacturing Survey produced by the *Dirección Administrativa Nacional de Estadística* or DANE (National Administrative Department of Statistics). This dataset records different outcomes for all manufacturing establishments than employ more than 10 workers or have revenues above a specific cutoff defined by DANE. THe cutoffs are updated every year using the Producer Price Index. The survey has been extensively used in previous international economics literature papers (Eaton et al, 2007; Eslava et al, 2013; Fieler et al, 2018). The producer price index for the manufacturing sector comes from the *Banco de la Republica* (the Central Bank of Colombia).

2.5. Empirical model

The objective of this paper is to analyze the impact of oil production in different manufacturing outcomes at a regional level in the short-run. There are three challenges for this: lack of cross-sectional and time variation of a commodity boom, availability of manufacturing data that includes detailed outcomes at a establishment level, and identification. As it was described in section 3, the oil boom in Colombia has enough cross-sectional and time variation for my empirical analysis. Moreover, as section 4 shows, detailed data regarding the manufacturing outcomes is available. The last issue, the identification of the elasticities of manufacturing outcomes with respect to oil production is addressed in this section.

The empirical model I use is the following:

$$X_{dt} = \alpha_d + \gamma_t + \beta\ (Oil\ production)_{dt} + \epsilon_{dt},\ where$$

X_{dt} represents different manufacturing sector outcomes for a department d in specific year t such as output, investment, payments to labor, etcetera; α_d is for department fixed effects, and γ_t represent time fixed effects. I run these regressions in logs, unless otherwise indicated. To identify the parameter of interest β, which represents the elasticity of a manufacturing outcome with respect to oil production, it is necessary that the unobservable ϵ_{dt} is uncorrelated with $(Oil\ production)_{dt}$, that is $E[\epsilon_{dt}\ (Oil\ production)_{dt}]$.

I focus on the short-run effects of oil production for two reasons. First, I follow the standard of the *resource curse* literature. Second, I cannot focus on the medium run given that in this time horizon domestic trade costs and input-output linkages can change substantially (e.g. input-output matrices change every 10 or 15 years). Third, there is existing work proving evidence of the medium and long-run spatial effects, but no previous work on the short-run spatial impacts.

The exogeneity assumption can be justified based on the Energy Economics literature. Anderson, Kellogg and Salant (2018) have shown that the main driver of oil production in fields is the *reservoir pressure*, that is, the hydrostatic pressure generated by fluids present in the oil field. As the oil is extracted from a field, this pressure slowly goes down, which generates lower oil production. Hence, oil production in a field is mainly governed by Darcy's Law, an equation that represent how a liquid flows through porous surfaces (Mason and Van't Veld, 2013). Hence, oil prices cannot influence oil production across fields. If this is the case, then it is very unlikely that local time varying unobservables can influence oil production.

A potential concern is that unobservable local conditions impact drilling activity, which itself has the potential to affect oil production. But both the economics literature and energy reports provide evidence that drilling activity might not be contemporaneously correlated with oil production. According to the economics literature, the drilling decisions are mainly influenced by three factors: geological characteristics, oil price levels, and oil price volatility (Anderson, Salant, and Kellogg, 2018; Kellogg, 2011). Nevertheless, these factors are captured in the department and time fixed effects. In addition, there are two relevant issues about drilling activity. First, after the finding of oil, at least one or two years can pass until the well starts producing oil, while in some cases the oil production can start five years later (see the brief report of Lieskovsky and Yan, 2019 about the North Dakota fields as an example). Second, finding oil after drilling is a probabilistic event. Even if oil prices influence drilling activity in a specific area, finding oil is random. Hence, it is safe to assume that even if unobserved local economic conditions influence drilling activity, it is unlikely that drilling activity influences oil production contemporaneously.

2.6. Results

In this section I provide evidence of the impact of oil production in the manufacturing outcomes at an aggregate level, and from a sectoral standpoint. I focus on the nominal values since I do not have access to local producer price indices. I present results in real terms too, but given the lack of local producer price indices, the results using real values need to be considered with caution.

2.6.1. Aggregate impacts on the manufacturing outcomes

The following tables 2.1, 2.2 and 2.3 provide estimates of the elasticities of different elasticities of output outcomes, labor outcomes and capital outcomes with respect to oil production, respectively. Results indicate oil production has little or no impacts on output, average wages. But it seems manufacturing establishments are increasing the capital stock as a consequence of the oil boom. Additionally, prices seem to be positively impacted by oil production since nominal output gross, but real output does not seem to change.

The previous results are in line with regions having strong industrial linkages and high domestic trade costs, as in scenario 2c in section 2. Previous work on transportation and economic history shows that Colombia has experienced historically high levels of domestic trade costs (Duranton, 2015; Pachon and Ramirez, 2007). In Section 6.2, I provide evidence that the oil boom benefits some industries but not others. Finally results at a sectoral and aggregate level indicate the number of manufacturing of establishments is completely unaffected by oil production. This suggests the results do not seem to be influenced by large increases in the demand of goods by households that migrate to the oil Colombian regions.

Output outcomes. I analyze the elasticities of gross production and industrial production[2] with respect to oil production. Notice an increase of 1 percentage point in the oil production leads to increases of 0.03% in the gross production and 0.02% in the indus-

[2]Gross production includes the production of goods that are used by the establishments themselves

trial production, measured in nominal values. The elasticities measured in real value are not statistically different than zero. Therefore, the results provide evidence that local oil production lead to increases in prices, and no changes in real output.

Table 2.1 Estimates of elasticities of variables related to output (in nominal values) with respect to oil production

	Dependent variable (in logs)			
	Gross production (Nominal value)	Industrial production (Nominal value)	Gross production (Real value)	Industrial production (Real value)
log(oil production)	0.03**	0.02**	0.01	0.13
	(0.01)	(0.01)	(0.01)	(0.14)
Observations	480	480	480	478
R-squared	0.42	0.42	0.27	0.07
Number of departments	32	32	32	32

Real values are in Colombian pesos of 2014
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Labor outcomes. I highlight three results related to labor outcomes. Firstly, the number of total workers in the manufacturing sector grows and the nominal total payments to labor also grow slightly. Secondly, the impact of oil production on average manufacturing wage (in both nominal and real terms) in a department is not statistically different than zero. Lastly, the total payments to labor in real terms does not seem to be affected by oil production.

Table 2.2 Estimates of elasticities of labor outcomes in nominal values with respect to oil production

	Dependent variables (in logs)				
	Total workers	Total payments to labor (Nominal value)	State average wage (Nominal value)	Total payments to labor (Real value)	State average wage (Real value)
log(oil production)	0.01*	0.02**	0.01	0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	480	480	480	480	480
R-squared	0.36	0.41	0.26	0.26	0.10
Number of departments	32	32	32	32	32

Real values are in Colombian pesos of 2014
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Capital and financial outcomes. Tables 2.3 and 2.4 show how regional oil production impacts the outcomes related to capital and financial outcomes of the firm, in nominal and real values respectively. The results in both tables show that manufacturing firms seem to be accumulating capital. It is noticeable that the rent of machinery is negatively impacted by the presence of oil production in the regions.

Demand by new households in oil regions. Given the lack of data regarding population or number of households in every department per year, I evaluate this mechanism

Table 2.3 Estimates of elasticities of capital and financial outcomes in nominal values, with respect to oil production

	Dependent variables (in logs)					
	Gross investment	Fixed assets	Rent of machines	Rent of buildings	Use of leasing for capital acquisition	Financial expenditures
log(oil production)	0.35**	0.05***	-0.21***	0.02	1.32***	-0.00
	(0.17)	(0.02)	(0.05)	(0.02)	(0.42)	(0.04)
Observations	480	480	480	480	224	352
R-squared	0.11	0.26	0.08	0.29	0.61	0.07
Number of departments	32	32	32	32	32	32

Real values are in Colombian pesos of 2014
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2.4 Estimates of elasticities of capital and financial outcomes in nominal values, with respect to oil production

	Dependent variables (in logs)					
	Gross investment	Fixed assets	Rent of machines	Rent of buildings	Use of leasing for capital acquisition	Financial expenditures
log(oil production)	0.13	0.05***	-0.22***	0.01	1.31***	-0.02
	(0.14)	(0.02)	(0.05)	(0.02)	(0.41)	(0.04)
Observations						
R-squared	478	480	480	480	224	352
Number of	0.07	0.26	0.08	0.18	0.61	0.05
departments	32	32	32	32	32	32

Real values are in Colombian pesos of 2014
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

by analyzing whether the number of establishments is impacted by oil production. Table 2.5 shows that oil production levels do not seem to influence the number of establishments. The estimates of the elasticities are very small and statistically not different than zero. The results indicate that commodity booms do not seem to cause large shifts in the demand of manufacturing goods by new households, which would lead to new establishments.

Table 2.5 Oil production and number of establishments

	Dependent variables: log (number of establishments)					
	All sectors	Food, beverage and tobaccco	Paper, textiles and leather	Chemical industries	Electrical machinery transportation	Other industries
log(oil production)	0.01	0.01	0.00	-0.02	-0.01	-0.01
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Observations	480	480	480	480	480	480
R-squared	0.21	0.09	0.06	0.15	0.07	0.16
Number of observations	32	32	32	32	32	32

Standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Intermediate inputs. Table 2.6 provides information on how regional oil production impacted the outcomes related to intermediate inputs. A regional oil boom does impact positively the value of inputs used, and the purchases of foreign inputs. On the other hand, the impact on the expenditure on transportation is not statistically different than zero. The results for intermediate inputs are consistent with the fact that the oil boom leads to increases in prices and output.

Table 2.6 Estimates of the elasticities of expenditure on intermediate inputs with respect to oil production

	Nominal variables (in logs)			Real variables (logs)		
	Value of inputs used	Expenditure on transportation	Purchase of foreign inputs	Value of inputs used	Expenditure on transportation	Purchases of foreign inputs
log(oil production)	0.04***	-0.00	0.37***	0.02*	-0.01	0.35***
	(0.01)	(0.03)	(0.12)	(0.01)	(0.03)	(0.12)
Observations	480	480	480	480	480	480
R-squared	0.45	0.05	0.05	0.31	0.34	0.05
Number of departments	32	32	32	32	32	32

Real values are in Colombian pesos of 2014
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

2.6.2. Impacts on manufacturing outcomes by sector-region

As mentioned in subsection 6.1, the results at an aggregate level are in line with a story related to IO linkages and high domestic trade costs. To provide stronger evidence of this, I analyzed how the oil production influences the previous outcomes at a sectoral-regional level. The results provide evidence that two groups of industries benefit from a regional oil boom: "Chemical Industries" and "Food, Beverages and Tobacco". The first group has strong backward and forward linkages the the oil sector (e.g. crude oil is necessary for the production of several chemicals, and crude oil also requires several chemicals as inputs). In addition, data from the input-output table shows that a large share of the expenditures of the oil sector is on goods produced by the food manufacturing sector.

To provide evidence of this, tables 2.7 and 2.8 provide information about the main activities that have backward and forward linkages with the economic activity "Oil and natural gas extraction, extraction of uranium and torium" according the input-output table of the official statistical agency of Colombia, DANE. It is noticeable that the activity "Oil and natural gas extraction, mining of torium and uranium" main purchases are from activities that can be classified in the categories "Food, beverages and tobbaco" and "Chemical industries" (backward linkages). Moreover, an activity that belongs to "Chemical industries" (Products from coal ovens; products from oil refining and nuclear fuel) benefits extensively from higher oil production, given that such activity uses crude oil as an input (forward linkages).

It is important to highlight that the impacts of oil production on different sectors also depend on how easy or difficult is the international logistics of the goods in these

=

Table 2.7 Backward linkages of the economic activity "Oil and natural gas extraction, extraction of uranium and torium"

Intermediate consumption of by the activity "Oil and natural gas extraction, mining of torium and uranium"	**%**
Meat products, frozen fish, prepared fish, canned seafood products.	35.5
Products from coal ovens; products from oil refining and nuclear fuel	15.0
Transportation services (cargo trucks and ducts)	32.2
Administrative services	5.4
Electricity	2.0

Table 2.8 Forward linkages of the economic activity "Oil and natural gas extraction, extraction of uranium and torium"

Activities that purchase goods from the activity "Oil and natural gas extraction, mining of torium and uranium"	**%**
Products from coal ovens; products from oil refining and nuclear fuel	76.1
Gas distribution services via ducts	15.6
Electricity generation	6.1
Professional , technical and research services	1.1
Basic chemical manufacturing, plaguicides, and fertilizers	1.0

sectors. In these sense there is another characteristic shared by the products in the categories "Food, beverages and tobacco" and "Chemical industries": they can difficult to transport. Food products in some cases have an expiration date (even if they are subject to chemical processing), some of them require refrigeration and many times they are subject to strict regulations. Many chemical products require special conditions for transportation: containers made of specific materials, products can react to changes in pressure, humidity or temperature, they require special inspections.

I report the results for nominal variables (the results for real variables can be consulted in Appendix B). I focus on these results due to the lack of regional price indices, hence the results in nominal value are more precise. Regarding the groups of industries, I created five categories. The categories are the following: "Food, tobbaco and beverages", "Wood, paper, leather, textiles (basic manufacturing)", "Chemical industries", "Electrical, mechanical and transportation", "Other basic manufacturing industries". I created broad categories to guarantee there were enough observations in most departments, but taking into consideration that these industries have common characteristics. For example, I created the group "Wood, paper, textiles, leather, since the technological needs of these sectors is low, and these industries are labor intensive.

Output outcomes. The estimates shown in table 2.9 show that the impacts of oil production are heterogeneous across sectors. That is, an increase of 1% in the oil production leads to increases of 0.76% and 1% in the nominal value of industrial production of "Food, Beverages and Tobacco" and "Chemical Industries", respectively. In an opposite way, a 1% growth in the oil production causes a decrease in the nominal value of industrial production of 1.22% and 1.06% for the case of "Wood, Paper, Textiles and Leather" and "Electrical, Machinery and Transportation Devices".

Table 2.9 Estimates of the elasticities of industrial production (nominal values) with respect to oil production

Dependent variable: industrial production (nominal values)					
	Food beverage tobaccco	Wood, paper textiles leather	Chemical sector	Electrical machinery transportation	Other industries
log(oil production)	0.77***	-1.20***	1.00***	-1.05***	0.10
	(0.26)	(0.23)	(0.20)	(0.19)	(0.08)
Observations	480	480	480	480	480
R-squared	0.69	0.37	0.23	0.17	0.05
Number of departments	32	32	32	32	32

Standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Wages. The estimates of the elasticities of nominal wages with respect to oil production are informative regarding the heterogeneous impacts across sectors (Table 2.10). Those groups of industries that experienced a growth in the value of industrial production also experience a growth in the average wages at a department level. Similarly, the groups of industries that experience a negative effect in the industrial production also see the wages go down.

Table 2.10 Estimates of the elasticities of nominal wages with respect to oil production

Dependent variable: average wage by department (nominal values)					
	Food beverage tobaccco	Wood, paper textiles leather	Chemical sector	Electrical machinery transportation	Other industries
log(oil production)	0.44***	-0.61***	0.61***	-0.63***	0.02
	(0.13)	(0.13)	(0.12)	(0.12)	(0.04)
Observations	480	480	480	480	480
R-squared	0.69	0.29	0.20	0.12	0.07
Number of departments	32	32	32	32	32

Standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Investment. The regional oil production impacts investment and fixed assets, similarly as the way it impact industrial output and wages (see Table 2.11). Therefore, the regional oil production not only seem to affect current production decisions, but also future production decisions (via capital acquisition). This implies that establishments expand or contract their capacity due to changes in the regional oil production.

Table 2.11 Estimates of the elasticities of gross investment (nominal terms) with respect to oil production

	Dependent variable: gross investment (nominal values)				
	Food beverage tobaccco	Wood, paper textiles leather	Chemical sector	Electrical machinery transportation	Other industries
log(oil production)	0.66***	-0.69**	0.72***	-0.32	-0.26
	(0.24)	(0.27)	(0.19)	(0.23)	(0.27)
Observations	480	477	479	479	480
R-squared	0.66	0.24	0.19	0.08	0.06
Number of departments	32	32	32	32	32

Standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 2.12 Estimates of the elasticities of value of fixed assets (in real terms) with respect to oil production

	Dependent variable: fixed assets (nominal values)				
	Food beverage tobaccco	Wood, paper textiles leather	Chemical sector	Electrical machinery transportation	Other industries
log(oil production)	0.81***	-1.18***	0.98***	-1.03***	0.02
	(0.26)	(0.22)	(0.20)	(0.18)	(0.08)
Observations	480	480	480	480	480
R-squared	0.69	0.36	0.23	0.17	0.03
Number of departments	32	32	32	32	32

Standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Intermediate inputs. In tables 2.13, 2.14 and 2.15, I report the impact of the oil production in three relevant outcomes regarding intermediate inputs: value of inputs used, purchases of foreign inputs and expenditures on transportation of intermediate inputs. I report the analysis for nominal values, while the results for real variables are included in Appendix B.

When I consider the value of intermediate inputs used and the expenditures of transportation of intermediate inputs, the results are in line with the impact of oil production in industrial production: the groups of industries that experience output growth ("Food, beverage and tobbacco" and "Chemical industries") also experience increases in the expenditures of the value of intermediate inputs and the transportation of these inputs. On the other side, the industries whose output is negatively impacted ("Wood, paper, textiles and leather" and "Electrical, machinery and transportation") reduce the expenditures on these two factors. The estimates for "Other industries" are not statistically different than zero.

Lastly, results regarding the impact of oil production in the purchases of foreign inputs are similar to the previous two variables, but for three categories, the estimates are not statistically significant: "Chemical industries", "Electrical, machinery and transportation", and "Other industries".

Table 2.13 Estimates of the elasticities of value of intermediate inputs used with respect to oil production

Dependent variable: value of intermediate inputs used (nominal values)					
	Food beverage tobaccco	Wood, paper textiles leather	Chemical sector	Electrical machinery transportation	Other industries
log(oil production)	0.74***	-1.15***	0.96***	-0.99***	0.09
	(0.26)	(0.22)	(0.20)	(0.18)	(0.08)
Observations	480	480	480	480	480
R-squared	0.69	0.38	0.24	0.16	0.05
Number of departments	32	32	32	32	32

Standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 2.14 Estimates of the elasticities of value purchases of foreign inputs used with respect to oil production

Dependent variable: purchases of foreign inputs (nominal values)					
	Food beverage tobaccco	Wood, paper textiles leather	Chemical sector	Electrical machinery transportation	Other industries
log(oil production)	0.85***	-1.04***	0.08	-0.18	0.01
	(0.27)	(0.20)	(0.16)	(0.14)	(0.19)
Observations	480	480	480	480	480
R-squared	0.35	0.13	0.04	0.05	0.09
Number of departments	32	32	32	32	32

Standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 2.15 Estimates of the elasticities of expenditures on transportation of intermediate inputs with respect to oil production

Dependent variable: expenditures on transportation of inputs (nominal)					
	Food beverage tobaccco	Wood, paper textiles leather	Chemical sector	Electrical machinery transportation	Other industries
log(oil production)	0.60***	-0.99***	0.75***	-0.69***	0.12
	(0.20)	(0.17)	(0.15)	(0.14)	(0.09)
Observations	480	480	480	480	480
R-squared	0.69	0.42	0.21	0.23	0.19
Number of departments	32	32	32	32	32

Standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

2.7. Conclusion

There are three main conclusions from this paper. Firstly, commodity booms have the potential to generate positive regional spillovers in the manufacturing sector in the short-run. This idea differs from previous economics work on commodity booms, which usually associates these commodity booms with negative short-run effects on manufacturing and political economy outcomes. Secondly, there are three factors that are relevant to determine the spatial effects of commodity booms: input-output linkages, domestic trade costs and complexity of the logistical transportation of the products. The difficulty in the logistical transportation of goods operates in the following way: sectors for which the transportation of goods requires special logistics tend to be produced locally, therefore they are also more likely to be impacted by local shocks. Third, the impacts of commodity booms in a region strongly depend on factors related to general equilibrium effects such as industrial linkages and domestic trade costs. Notice that the previous factors are different than the medium and long-run spatial impacts of commodity booms (e.g. agglomeration) that are studied in previous literature.

Specifically, in the context of Colombia, I provide evidence that the impacts of commodity booms strongly depend on the input-output linkages and the high domestic trade costs in the country. An oil commodity boom leads to higher manufacturing output in regions where the chemical manufacturing and the food, beverages and tobacco industries have a higher presence. Moreover, the transportation of goods for these sectors has higher requirements due to their chemical and physical characteristics, hence local demand would tend to be more sensitive to changes in oil production. Differently, regions with a strong presence of the manufacturing of textiles, leather, wood and paper, as well as establishments that manufacture mechanical, electrical and transportation devices would see their manufacturing output negatively impacted. Interestingly, most of the products of these sectors are easier to transport (relative to chemical or food products).

The ideas in this paper are highly relevant for the vast existing economics literature on the *resource curse*. Analyzing the short-run regional outcomes on labor markets, political economy variables, public finance, and development indicators across regions in a country without taking into consideration the economic geography effects of commodity booms can lead to biased conclusions. For example, the regional impacts of commodity booms are not the same in regions with very low domestic trade costs vs. regions with high domestic trade costs. Similarly, the manufacturing composition of the region matters substantially when. Therefore, by ignoring the spatial and geographical factors of commodity booms, the conclusions of the resource curse literature lack external validity.

Lastly, this paper shows the relevance of IO linkages and domestic trade costs when we evaluate large regional shocks that have spatial effects. Commodity booms are not the only economic phenomena that lead to substantial regional impacts. Large migration waves, climate change events or big increases in greenfield foreign direct investment can have heterogeneous regional impacts that depend on IO linkages and domestic trade costs. Therefore, existing applied microeconomics literature needs to take into consideration these factors, before making conclusions regarding how such events impact regional economic variables.

2.8. Bibliography

1. Anderson, S., Kellogg, R., and Salant, S. (2018). *Hotelling Under Pressure.* Journal of Political Economy. Vol. 126, No. 3, 984-1026.

2. Allscott, H. and D. Keniston (2018). *Dutch Disease or Agglomeration? The Local Economic Effects of Natural Resource Booms in Modern America.* Review of Economic Studies, Vol. 85, No. 2, 596-731.

3. Antras P. and A. De Gortari (2020). *On the Geography of Gloval Value Chains*, Econometrica, Forthcoming.

4. Aragon, F.M. and Rud J.P. (2013). *Natural Resources and Local Communities: Evidence from a Peruvian Gold Mine.* American Economic Journal: Economic Policy, Vol. 5, Issue 2, 1-25.

5. Aragon, F.M. and Rud J.P. (2016). *Polluting Industries and Agricultural Productivity: Evidence from Mining in Ghana.* Economic Journal, Vol. 126, Issue 597, 1980-2011.

6. Arezki, R., Ramey, V., and Sheng L. (2017). *News Shocks in Open Economies: Evidence from Giant Oil Discoveries.* The Quarterly Journal of Economics, Vol. 132, Issue 1, 103-155.

7. Arezki, R., and M. Brückner (2012). *Commodity Windfalls, Democracy and External Debt.* The Economic Journal, Vol. 122, Issue 561, 848-866.

8. Balashundaram, V. (2019). *Bottlenecks versus ripple effects: The role of linkages in India's product market liberalization.* Mimeo, University of Michigan.

9. Baqaee, D. (2018). *Cascading Failures in Production Networks*, Econometrica, Vol. 86, Issue 5, 1819-1838.

10. Baqaee, D. R. and E. Farhi (2019). *The Macroeconomic Impact of Microeconomic Shocks: Beyond Hulten's Theorem.* Econometrica, Vol. 87, Issue 4, 1155-1203.

11. Barrot, J.N., and J. Sauvagnat (2016). *Input Specificity and the Propagation of Idiosyncratic Shocks in Production Networks*. The Quarterly Journal of Economics, Vol. 131, Issue 3, 1543-1592.

12. Boehm, C., Flaeen A., and N. Pandalai-Nayar (2019). *Input Linkages and the Transmission of Shocks: Firm-Level Evidence from the 2011 Tōhoku Earthquake*. The Review of Economics and Statistics, Vol. 101, Issue 1, 60-75.

13. Bjornland, H., L.A. Thorsrud, and R. Torvik. (2019). *Dutch Disease Dynamics Reconsidered*. European Economic Review, 119, 2019, 411-433.

14. Brollo, F., T. Nannicini, R. Perotti, and G. Tabellini (2013). *The Political Resource Curse*. American Economic Review, Vol. 103, No. 5, 1759-1796.

15. Bruno, M., and J. Sachs (1982). *Energy and Resource Allocation: A Dynamic Model of the "Dutch Disease"*. The Review of Economic Studies, Vol. 49, Issue 5, 845-859.

16. Caliendo, L. and F. Parro (2015). *Estimates of the Trade and Welfare Effects of Nafta*. The Review of Economic Studies, Vol. 82, No. 1, 1-44.

17. Carreri, M. and Dube, O. (2017). *Do Natural Resources Influence Who Comes to Power, and How?*. Journal of Politics, Vol. 79, Issue 2, 502-518.

18. Carvalho, V., M. Nirei, Y. Saito, and A. Tahbaz-Salehi (2016). *Supply Chain Disruptions: Evidence from the Great East Japan Earthquake*, Columbia Business School Research Paper No. 17-5

19. Caselli, F., and A. Tesei (2016). *Resource Windfalls, Political Regimes, and Political Stability*. Review of Economics and Statistics, Vol. 98, Issue 3, 573-590.

20. Caselli F., and G. Michaels (2013). *Do Oil Windfalls Improve Living Standards? Evidence from Brazil*. American Economic Journal: Applied Economics, Vol. 5, No. 1, 208-238.

21. Cassidy, T. (2019). *The Long-Run Effects of Oil Wealth on Development: Evidence from Petroleum Geology.* The Economic Journal. Vol. 129, 2745-2778.

22. Corden W.M., and J.P. Neary (2002). *Booming Sector and De-Industrialization in a Small Open Economy.* The Economic Journal, Vol. 92, No. 368, 825-848.

23. Costinot, A., J. Vogel, and S. Wang (2013). *An Elementary Theory of Global Supply Chains*, The Review of Economic Studies, Vol. 80, Issue 1, 109-144.

24. Dube, O., and J.F. Vargas (2013). *Commodity price shocks and civil conflict: Evidence from Colombia.* Review of Economic Studies, Vol. 80, Issue 4, 1384-1421.

25. Eaton, J., M. Eslava, M. Kugler, and J. Tybout (2007). *Export Dynamics in Colombia: Firm-Level Evidence*, Borradores de Economía, 3957, Banco de la República.

26. Eslava M., J. Haltiwanger, A. Kugler and M. Kugler. *Trade and market selection: Evidence from manufacturing plants in Colombia.* Review of Economic Dynamics, Vol. 16, Issue 1, 135-158.

27. Feyrer, J., E. Mansur and B. Sacerdote (2017). *Geographic Dispersion of Economic Shocks: Evidence from the Fracking Revolution.* American Economic Review, Vol. 107, No. 4, 1313-1334.

28. Fieler A.C., M. Eslava and D.X. Yu (2018). *Trade, quality upgrading, and input likages: Theory and Evidence from Colombia.* American Economic Review, Vol. 108, Issue 1, 109-146.

29. Grassi, B. (2018). *IO in I-O: Competition and Volatility in Input-Output Networks.* Mimeo, Bocconi University.

30. Huo, Z., A. Levchenko, and N. Pandalai-Nayar (2020). *nternational Comovement in the Global Production Network*, NBER Working Paper 25978

31. Ishak, P.W. and P.G. Meon (2020). *A resource-rich neighbor is a misfortune: The spatial distribution of the resource curse in Brazil.* Working Paper CEB 20-001, Universite Libre de Bruxelles.

32. Kellogg, R. (2014). *Investment: Evidence from Texas Oil Drilling.* American Economic Review. Vol. 104, No. 6, 1698-1734

33. Litschig, S. (2012) *Are rules-based government programs shielded from special-interest politics? Evidence from revenue-sharing transfers in Brazil.* Journal of Public Economics, Vol. 96, Issues 11-12, 1047-1060.

34. Monteiro J. and Ferraz C. (2012). *Does Oil Make Leaders Unnacountable? Evidence from Brazil's offshore boom.* Mimeo, PUC-Rio.

35. Michaels, G. (2011). *The Long Term Consequences of Resource-Based Specialisation.* The Economic Journal, Vol. 121, Issue 551, 31-57.

36. Yi, Kei-Mu (2003). *Can Vertical Specialization Explain the Growth of World Trade.* Journal of Political Economy, Vol. 111, No. 1, 52-102.

Chapter 3. Dynamics of Importers after Trade Liberalization

3.1. Introduction

A large literature on trade liberalization has focused on the export behavior of firms: how they enter to specific markets, their dynamics, and how they react to large shocks. However, less is known about the behavior of the importers. Although we have some evidence on the response of importers to large devaluations, the pass through of changes in tariffs and the role of imports on firms' productivity, there is limited literature focusing on the dynamics of importers. The objective of this paper is to analyze the dynamic and heterogeneous behavior of importers, and how this behavior changes in response to a trade liberalization event. This would shed some light into the constraints for importing firms while improving our understanding of how the composition of importing firms and imports change post trade liberalization.

We use customs data from Colombia to study the heterogeneous response of Colombian imports to the Free Trade Agreement (FTA) between USA and Colombia in 2012. We show that the overall effects of FTA in the case of Colombia are consistent with experiences in other contexts. In line with existing quantitative evidence and theoretical predictions, importers are generally more productive establishments irrespective of the FTA. The FTA also resulted in the exit of establishments and almost all of the exit is from firms that exclusively purchased local inputs. This suggests that the FTA lead to re-alloacation in factors of production towards larger and more productive firms, likely resulting in overall productivity growth. Hence, the Colombian episode offers a trade liberalization study that would be externally valid.

We document three novel patterns among importers: churning of importers, convergence of new importers with respect to old (or existing) importers, and divergence among existing importers. First, we observe churning of importers every year. Specifically, we observe entry of new importers while old, small and lumpy importers stop their importing behavior. Second, we document convergence behavior among new importers with respect to old importers. Moreover the new importers seem to overtake the existing small and lumpy importers. The convergence of new importers and the churning of importers are related and can be explained by a very high growth among new importers. Third, we observe *divergence among existing importers*, that is, there is gap between "frequent and larger importers" and "small and lumpy importers".

We also document how these patterns change across time in the period 2008-2018, and whether a Free Trade Agreement between Colombia and the United States has an impact in these patterns. Our findings show that these behaviors are not affected by the free trade liberalization of Colombia in 2012. Additionally, we show that the new importers are existing firms that did not import prior to FTA and not large new entrants into the market. This alleviates concerns that a "maquiladora boom" might be driving the observed import behaviors.[1]

Our paper contributes to the existing literature in two ways. Firstly, there is a large literature that studies the difference between firms that trade and those that do not (Bernard, Jensen and Schott, 2005; Bernard, Jensen and Schott, 2009; Bernard, Redding and Schott, 2004; Melitz, 2003). Moreover, our knowledge regarding the dynamics of exporters is expansive (Albornoz et al, 2012; Eaton, Kortum and Kramarz, 2011; Gilbert, 2018; Morales, Scheu and Zahler, 2019). While there is substantial literature that studies the impact of imports on productivity (Amiti and Konings, 2007; Goldberg et al., 2010; Halpern et al.; 2015) or how devaluations impact the composition imports (Gopinath

[1] NAFTA generated a boom of Mexican establishments called "maquilas" or "maquiladoras". These establishments purchase American inputs and use local cheap labor to assemble products, and then export them back to the United States. One characteristic of these establishments is that the share of foreign inputs is extremely high.

and Neiman, 2014; Alessandria et al., 2010), there is limited literature that studies the dynamics of importers (Lu et al., 2017). By exploring heterogeneous importer behavior, we can shed some light into the channels through which the gains of trade and productivity impacts are realized.

The remainder of the paper proceeds as follows. Section 2 provides background on the FTA between Colombia and USA. Section 3 describes the data and the construction of the lumpiness and experience measures. Section 4 proceeds to present the identification strategy with tests of the main identifying assumptions. Section 5 analyzes the impact of the FTA on firm import patterns by firm characteristics and briefly discusses the effects at the industry-level. Section 6 concludes.

3.2. Colombia-USA Free Trade Agreement

Although Colombia experienced unilateral trade liberalization in 1991, the tariffs on American products remained hefty, with 55% of the products that were negotiated in the Colomobia-USA FTA having tariffs in the range of 10 to 20% (Ministerio de Comercio, 2020).

Colombia and the United States negotiated for almost 21 months for a bilateral Free Trade Agreement and signed the trade deal in February 2006. After this, the Colombian Congress approved the FTA in July 2007. Although George W. Bush sent to the American Congress the final agreement in 2006, disagreements between Republicans and Democrats lead to delays in its approval. It wasn't until October 2011 that both the House of Representative and the Senate approved the FTA. The trade agreement formally was put in place on May 2012.

Using data from the website of the Free Trade Agreement created by the Colombian Ministry of Commerce, Figure 3.1 shows the number of products by sector where the tariffs were reduced to 0%. For most sectors, the tariffs of more than half of the products were completely eliminated. The FTA lead to an initial boom in the American imports between 2012 and 2015, as it can be seen in figure 3.2. By 2019 the flow of imports consistently remained higher than the years prior to the start of the trade agreement.

Figure 3.1 Percentage of products negotiated in the FTA with zero tariffs after the trade deal started.

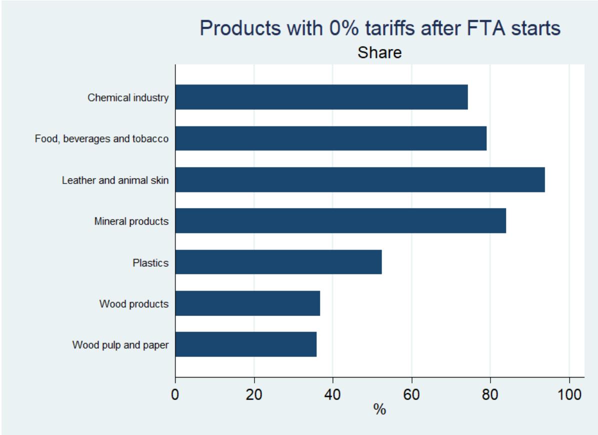
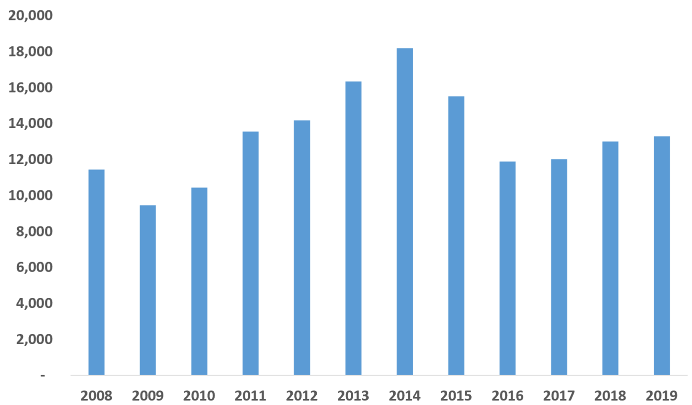


Figure 3.2 Colombian imports from the USA (millions of dollars)



3.3. Data

We use two different datasets for our analysis. The first dataset includes information about the Colombian manufacturing establishments. The second, is a comprehensive dataset of all the imports of Colombia. We focus on the period 2008 to 2011 to avoid capturing the effects of the financial crisis of 2008. The first dataset is every Annual Manufacturing Survey for the above period. This dataset produced by the official statistical institute of Colombia, the *Direccion Administrativa Nacional de Estadística* or DANE (National Administrative Directorate of Statistics). This manufacturing survey includes all establishments with more than 10 workers, or with revenues above a specific cutoff, which is updated using the producer price index every year. Given the design of the survey, this is equivalent to a census of manufacturing establishments. The dataset is very comprehensive and includes extensive detailed information about output, labor, capital, financial expenditures, and intermediate inputs.

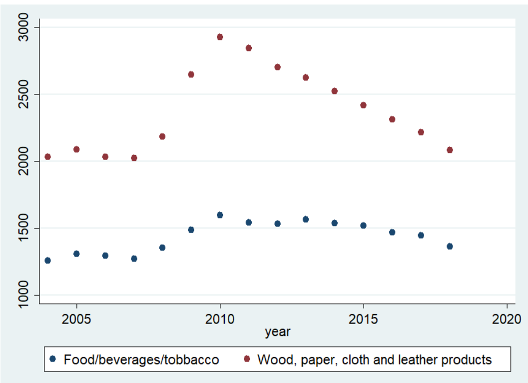
The second dataset is the customs administrative data produced by the tax authority of Colombia, the *Direccion de Impuestos y Aduanas Nacionales* or DIAN (Directorate of Taxes and National Customs). Specifically, we use the import data from 2008 to 2018. This dataset is fairly detailed and includes information regarding the tax ID of the firm, country of origin, HS 10-digit code, value in customs, insurance, expenditures on transportation, tariff, method of transportation, etcetera. Given that we have the tax ID of the firm, it is possible for us to capture the individual imports of every firm. Since part of the fiscal revenue of Colombia still depends on tariffs, this information is reliable since it is in the best interest of the Colombian government to generate reliable data.

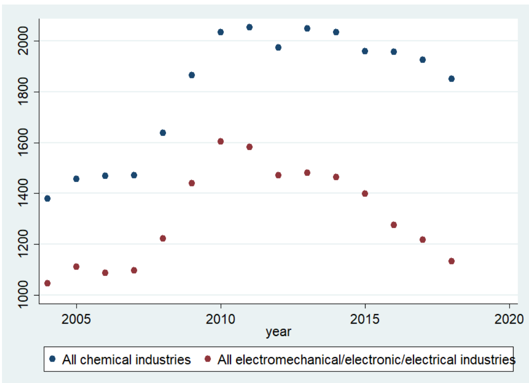
3.4. Empirical facts about Colombian establishments

In this section, we document some empirical patterns regarding the behavior of establishments after the free trade liberalization in Colombia.

Fact 1. A large exit of firms occurs after the FTA starts. We find a large exit of firms post the implementation of FTA. This is in line with increased competition driving out less productive firms, as Melitz (2003) predicts. While this is consistent across different industries, it is particularly stark in textiles and electronic industries.

Figure 3.3 Number of manufacturing establishments across years

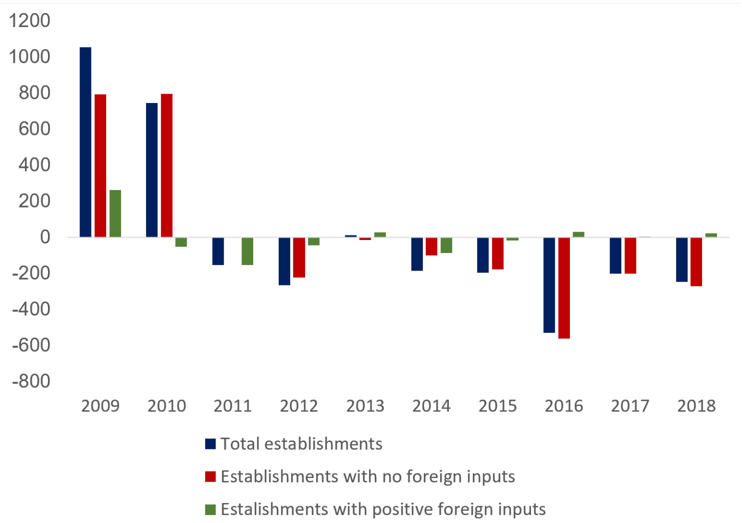




Note: I created these broad categories by grouping 2-digit ISIC manufacturing sectors. See Appendix A for more information.

Fact 2. Most of the exit of firms occurs among non-importers. Using the census data for manufacturers, we show that most of the exit of establishments occurs across non-importers. More importantly, between 2011 and 2018 very few establishments with positive foreign inputs exit.

Figure 3.4 Changes in the number of establishments classified according to whether they import or not



Fact 3. Only the most productive firms import and this does not change after the FTA. Using the census of manufacturers, we show the evolution of total factor productivity measures (TFPQ) of the establishments that import and those who do not purchase foreign inputs using a generalized version of the Hsieh and Klenow (2009) methodology. Two conclusions can be drawn from the figure 3.5. Firstly, the productivity of importers is always higher than the productivity of non-importers. Secondly, the FTA does not change this gap.

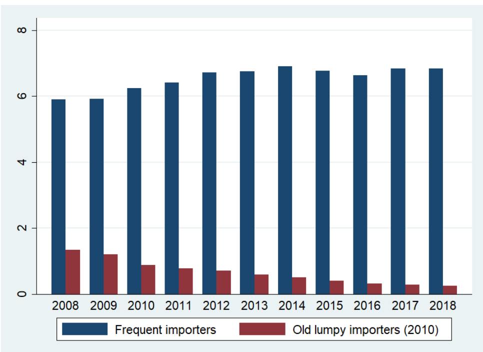
Figure 3.5 Evolution of TFPQ among manufacturing establishments

3.5. Heterogeneity of importers: descriptive analysis

This section briefly documents the existence of heterogeneous importers. To do so, we compare two types of existing importers: old lumpy importers versus large and frequent importers. We generate a cutoff τ for the years 2010, 2012, 2014 and 2016. The cutoffs allow us to define existing importers. For example, for the cutoff $\tau = 2010$, the existing importers are those that already imported before this year (e.g. in the years 2008 and 2009). The old lumpy importers are firms that imported before the cutoff, but did not import every year. The category frequent importers includes firms that imported every year between 2008 and 2018 (given that they imported constantly, they are considered existing importers). For all the cutoffs, we document two variables: the average value of imports and the average number of country-suppliers.

The following figures 3.6, 3.7, 3.8 and 3.9 provide evidence of the heterogeneous behavior of importers, independently of the cutoff selected. While the frequent importers increase constantly the average number of country suppliers, and experience a temporal growth in the size of the average size of imports, the old lumpy importers reduce both the average number of country suppliers and the mean value of the imports. Finally, there is no strong evidence that the FTA changes this heterogeneity of behavior among existing importers.

Figure 3.6 Behavior of lumpy importers vs. frequent importers when the cutoff is $\tau = 2010$ (left: number of country suppliers; right: average imports)



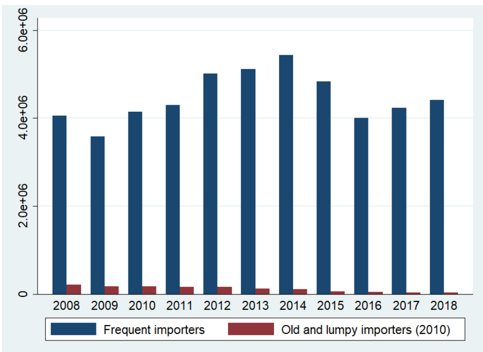


Figure 3.7 Behavior of lumpy importers vs. frequent importers when the cutoff is $\tau = 2012$ (left: number of country suppliers; right: average imports)

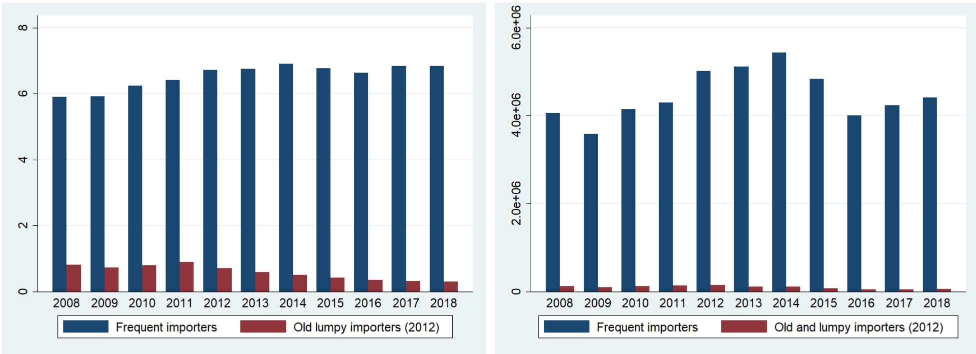


Figure 3.8 Behavior of lumpy importers vs. frequent importers when the cutoff is $\tau = 2014$ (left: number of country suppliers; right: average imports)

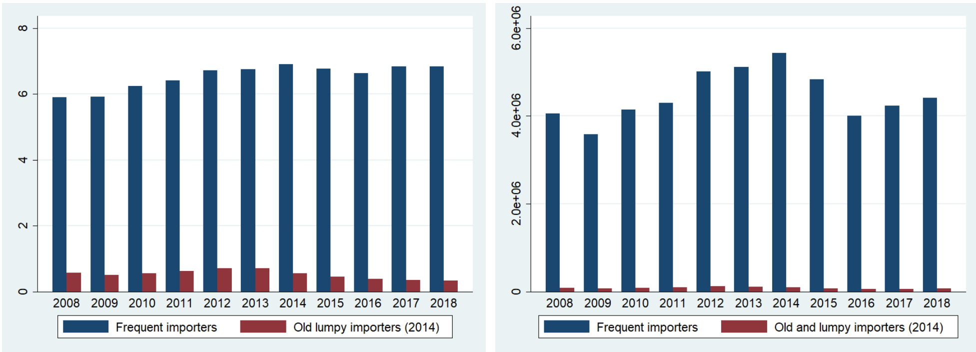
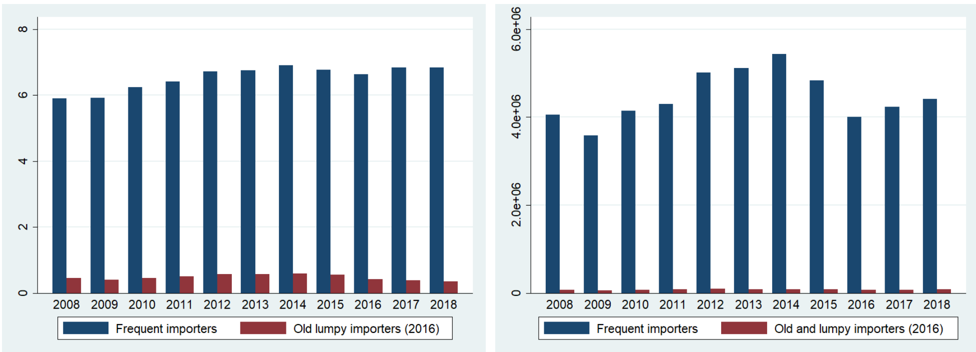


Figure 3.9 Behavior of lumpy importers vs. frequent importers importers when the cutoff is $\tau = 2016$ (left: number of country suppliers; right: average imports)



3.6. Churning of importers: descriptive analysis

We document in this section the churning of importers. This is the continuous exit of old lumpy importers, and a constant entry of new importers. Moreover, the new importers seem to overtake the old lumpy importers in the intensive and extensive margins of trade. More specifically, the new importers eventually have a higher average value of imports and more country suppliers than old lumpy importers.

Similar to the previous section, we generate cutoffs for the years 2010, 2012, 2014, and 2016. For example, we define a new importer in 2010, as an importer that reported zero importing activity in the years 2008 and 2009. We define an old and lumpy importer as a firm that had positive imports before the cutoff, but it did not import every year in the period 2008-2018.

The figures 3.10, 3.11, 3.12 and 3.13 provide evidence that the churning of importers occurs frequently, and it is not impacted by the trade liberalization. Interestingly, in the cutoffs of $\tau = 2014$ and $\tau = 2016$, the new importers converge with the old and lumpy importers much faster than in previous years.

Figure 3.10 Churning of importers when the cutoff is $\tau = 2010$ (left: average number of country suppliers, right: average size of imports)

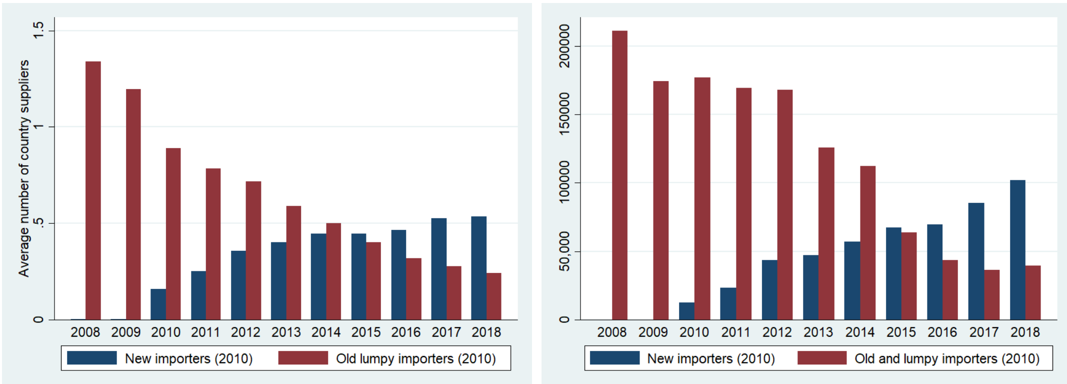


Figure 3.11 Churning of importers when the cutoff is $\tau = 2012$ (left: average number of country suppliers, right: average size of imports)

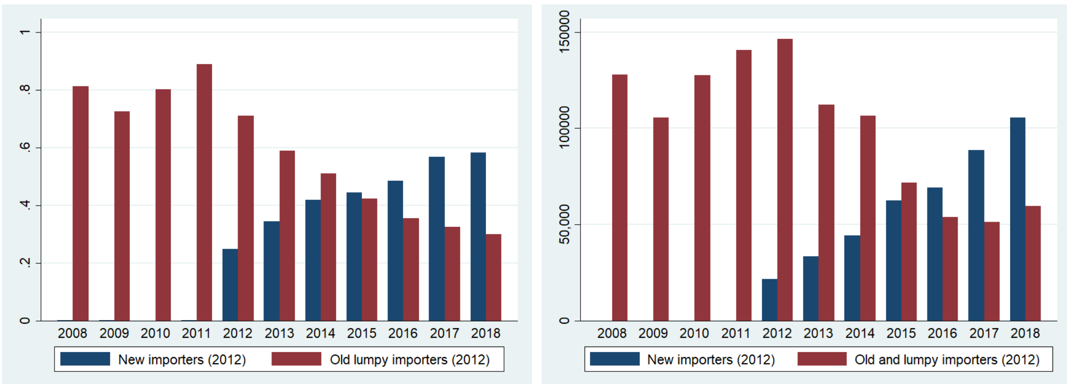


Figure 3.12 Churning of importers when the cutoff is $\tau = 2014$ (left: average number of country suppliers, right: average size of imports)

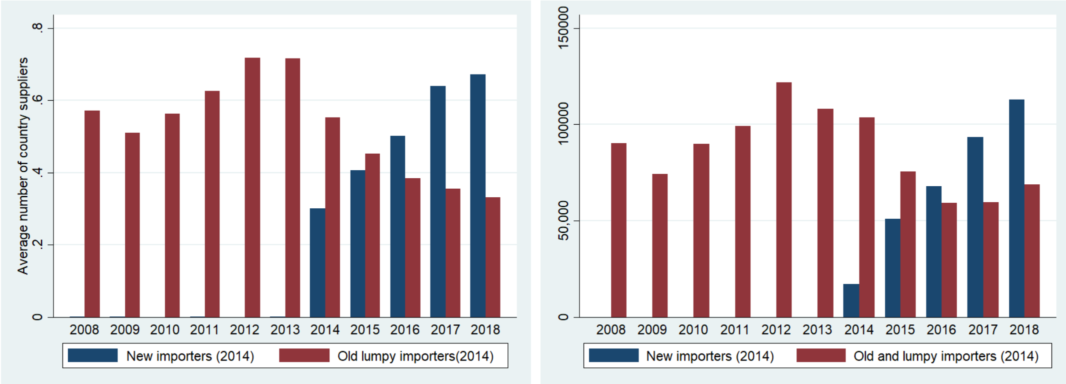
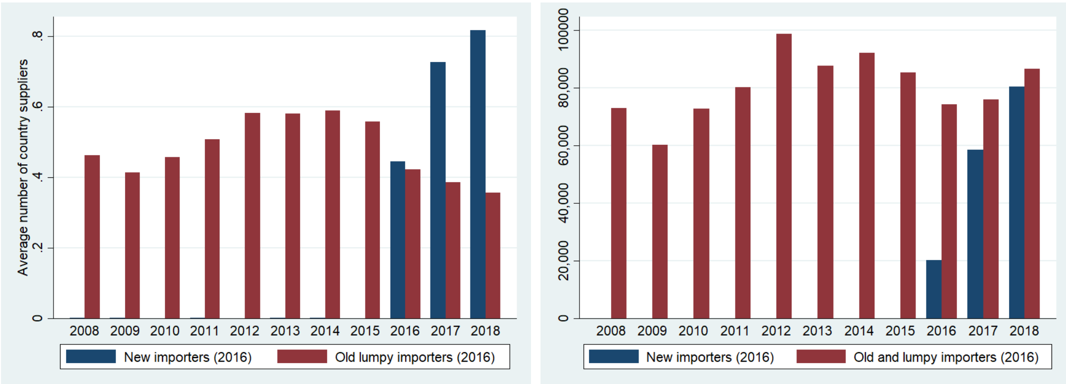


Figure 3.13 Churning of importers when the cutoff is $\tau = 2016$ (left: average number of country suppliers, right: average size of imports)



3.7. Econometric model

In this section, we present the empirical specifications to quantify two factors. Firstly, the convergence of new importers that start their importing activity in year τ, with respect to existing importers. Secondly, the divergence among existing importers in year τ, that is, the gap between old lumpy and existing frequent importers. We use the following standard event-study specification:

$$Y_{it} = \alpha + f_t + c_i + \beta \cdot 1[t \geq \tau]_t +$$

$$\gamma \cdot 1[t \geq \tau]_t \cdot (Interactionterm) + \epsilon_{it} \qquad (1)$$

where i is the establishment and t is the year. $1[t \geq \tau]_t$ is an indicator that that equals 1 if the year t is after the year of interest τ. Notice that when $\tau = 2012$, γ represents the existence of heterogeneous behavior of importers after the FTA is implemented. We ran regressions for the different values of τ, specifically for all the years between 2009 and 2017. All regressions include establishment fixed effects and year fixed effects. Y_{it} includes all the output variables of interest: Import share from USA, Number of country-suppliers, Number of sectors the firm purchases inputs from at the 2-digit HS level and a dummy that equals 1 if the firm imports from USA in a specific year (in which case we use the linear probability model). Since we are interesting in understanding the heterogeneous response of firms, we also include an interaction term with the treatment indicator. The interaction looks at the experience and frequency of importers. Experience is measured by the fraction of years of non-zero imports between 2008 and year τ. Frequency is measured as a dummy that equals 1 if the firm imported in all years between 2008 and 2018. Errors are clustered at the firm level to adjust for heteroskedasticity and within-firm correlation over time. For the aggregate industry-level analysis, we run the same regression as above but with industry fixed effects and errors clustered at the industry-level.

The main issue with our empirical model is that we cannot distinguish between new firms and new importers. Particularly, we cannot identify if firms that start to import after FTA are new entrants into the market or existing firms who were on the margin prior to FTA and became importers after the FTA. Secondly, by construction, experienced and frequent importers are more likely to import than new firms. Hence, our estimates of γ are an upper bound. Nevertheless, since the estimates are negative when we compare experienced importers with new importers, this alleviates the issue. Finally, when we compare old lumpy vs. old frequent importers the estimates are not an upper bound, given that both have experience importing activity prior to the FTA.

3.8. Results

In this section, we present results regarding the heterogeneous behavior of exporters. Firstly, we focus in the convergence behavior between new importers and old or existing importers (this includes both the frequent and lumpy ones). Therefore, we want to evaluate if the new importers always present higher growth in the importing behavior, relative to the old ones. Secondly, we analyze the divergence among old or existing importers. That is, we test whether large and frequent importers with the small and lumpy importers show divergent behavior. We look at the following four variables: share of exports from the USA, probability of importing goods from the United States, number of country suppliers (we include a measure that controls for inventories), and number of sectors the importer purchases goods from.

3.8.1. Measuring the convergence between new and old importers

Our findings highlight two trends regarding the convergence between new and old importers. The first trend is that new importers converge with old importers on all the importing outcomes that we analyze. The second trend there is no evidence that the FTA alters the convergence patterns substantially.

Importing from the USA. We analyze the differences between old and new importers with respect to two variables: the share of American imports and the probability of importing goods from the USA. We evaluate the new importers of every year between 2009 and 2017, and we find that new importers always show a convergence behavior with respect to old importers in both measures.

It is noticeable that in the first three years, the convergence of the new importers is relatively high. Starting in 2012 and until 2018, the estimates related to the share of American imports are relatively similar (table 3.1). This implies that the new importers that enter after the implementation of NAFTA show the same convergent behavior, that the new importers in subsequent years. That is, the new importers of 2012 (the year the FTA starts) catch up with old importers with the same intensity as the new importers of the subsequent years. This is also true when we consider the probability of importing goods from the United States (table 3.2). Interestingly, the convergence of new importers of 2009 and 2010 with respect to existing ones is stronger than the convergence of new importers considered in subsequent years, once the free trade agreement was already implemented.

Table 3.1 Comparison between new and old importers regarding the share of American imports

	t = 2009	t = 2010	t = 2011	t = 2012 (start of FTA)	t = 2013	t = 2014	t = 2015	t = 2016	t = 2017
1[Year¿ t]* 1[Imported at least once before year t]	-0.25*** (0.00)	-0.17*** (0.00)	-0.14*** (0.00)	-0.11*** (0.00)	-0.11*** (0.00)	-0.10*** (0.00)	-0.10*** (0.00)	-0.11*** (0.00)	-0.13*** (0.00)
Observations	1,399,475	1,399,475	1,399,475	1,399,475	1,399,475	1,399,475	1,399,475	1,399,475	1,399,475
R-squared	0.03	0.03	0.03	0.02	0.03	0.02	0.02	0.02	0.01
Number of id	127,225	127,225	127,225	127,225	127,225	127,225	127,225	127,225	127,225

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3.2 Comparison between lumpy and frequent importers regarding the probability of importing from the USA

	t = 2009	t = 2010	t = 2011	t = 2012 (start of FTA)	t = 2013	t = 2014	t = 2015	t = 2016	t = 2017
1[Year¿ t]* 1[Imported at least once before year t]	-0.33*** (0.00)	-0.23*** (0.00)	-0.18*** (0.00)	-0.14*** (0.00)	-0.14*** (0.00)	-0.14*** (0.00)	-0.14*** (0.00)	-0.14*** (0.00)	-0.17*** (0.00)
Observations	1,399,475	1,399,475	1,399,475	1,399,475	1,399,475	1,399,475	1,399,475	1,399,475	1,399,475
R-squared	0.03	0.03	0.03	0.02	0.03	0.02	0.02	0.02	0.01
Number of id	127,225	127,225	127,225	127,225	127,225	127,225	127,225	127,225	127,225

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Number of country suppliers. We evaluate the convergence in the number of country suppliers (the number of countries importers buy goods from). The results of the different regressions are in table 3.3. Similarly to the previous results, the convergence behavior of new importers in the cutoff $\tau = 2012$ is very similar to the convergence behavior of the new importers that start their importing activity in the subsequent cutoffs. Interestingly, the new importers of 2009 and 2010 have a higher convergence behavior, relative to the new importers in the following years. As a robustness check, we adjust the number of country suppliers given that some importers accumulate inventories. To do so we follow the methodology of Lu et al (2017). The results are shown in table 3.4, and they lead to the same conclusions as in the case in which we do not consider the accumulation of inventories.

Table 3.3 Comparison between old and new importers regarding the number of country-suppliers

	t = 2009	t = 2010	t = 2011	t = 2012 (Start of FTA)	t = 2013	t = 2014	t = 2015	t = 2016	t = 2017
1[Year¿ t]* 1[Imported at least once before year t]	-1.26*** (0.0083)	-0.88*** (0.0056)	-0.74*** (0.0046)	-0.68*** (0.0042)	-0.65*** (0.0041)	-0.65*** (0.0042)	-0.70*** (0.0047)	-0.76*** (0.0056)	-0.90*** (0.0075)
Observations	1,399,475	1,399,475	1,399,475	1,399,475	1,399,475	1,399,475	1,399,475	1,399,475	1,399,475
R-squared	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01
Number of id	127,225	127,225	127,225	127,225	127,225	127,225	127,225	127,225	127,225

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3.4 Comparison between old and new importers regarding the number of country-suppliers (adjusting for inventory issues using Lu et al (2017) methodology)

	t = 2009	t = 2010	t = 2011	t = 2012 (Start of FTA)	t = 2013	t = 2014	t = 2015	t = 2016	t = 2017
1[Year¿ t]* 1[Imported at least once before year t]	-0.93*** (0.01)	-0.88*** (0.01)	-0.75*** (0.00)	-0.68*** (0.00)	-0.65*** (0.00)	-0.65*** (0.00)	-0.70*** (0.00)	-0.77*** (0.01)	-0.91*** (0.01)
Observations	1,397,385	1,399,332	1,399,332	1,399,332	1,399,332	1,399,332	1,399,332	1,399,332	1,399,332
R-squared	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
Number of id	127,035	127,212	127,212	127,212	127,212	127,212	127,212	127,212	127,212
Standard errors in parentheses									

*** p<0.01, ** p<0.05, * p<0.1

Number of importing sectors. We also test the convergence of new importers with respect to old importers using a different measure of the extensive margin, the number of sectors the importers purchase goods from. Table 3.4 provides estimates of the different regressions that evaluate the *catch up* behavior of importers.

Interestingly, the behavior of new importers of the cutoff $\tau = 2012$ is similar to the ones in the cutoffs $\tau = 2013$ and $\tau = 2014$. Different than in previous outcomes, the new importers of the cutoffs $\tau = 2009$ and $\tau = 2010$ have a similar convergence as the new importers of the cutoffs $\tau = 2016$ and $\tau = 2017$. Finally, the FTA did not change substantially the convergence of sector suppliers of the new importers with respect to the old importers.

Table 3.5 Comparison between old and new importers regarding the number of sector-suppliers

	t = 2009	t = 2010	t = 2011	t = 2012 (Start of FTA)	t = 2013	t = 2014	t = 2015	t = 2016	t = 2017
1[Year> t]* 1[Imported at least once before year t]	-1.95*** (0.01)	-1.38*** (0.01)	-1.18*** (0.01)	-1.07*** (0.01)	-1.03*** (0.01)	-1.05*** (0.01)	-1.13*** (0.01)	-1.26*** (0.01)	-1.49*** (0.01)
Observations	1,399,332	1,399,332	1,399,332	1,399,332	1,399,332	1,399,332	1,399,332	1,399,332	1,399,332
R-squared	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
Number of id	127,212	127,212	127,212	127,212	127,212	127,212	127,212	127,212	127,212
Standard errors in parentheses									
*** p<0.01, ** p<0.05, * p<0.1									

3.8.2. Measuring the divergence among existing importers

We measure the divergence among old or existing importers on different outcomes. Specifically, we compare the behavior of "frequent" and "lumpy" importers on the four outcomes mentioned above. Our results can be summarized in the following way. Firstly, we observe the existence of divergence among old importers on all variables. Secondly, before and after the FTA this divergence shows constant reductions, therefore there is no evidence that the FTA alter the evolution of the divergence among existing importers. Third, the fall in the divergence of old importers for the variables related to American imports is very strong, to the point that in 2016, such divergence seems to disappear.

Importing from the United States. The results regarding divergence among existing importers, on outcomes the share of American imports and probability of importing goods from the USA are shown in tables 3.6 and 3.7, respectively. Both tables provide evidence that the divergence among existing importers has a downward trend, and towards the end of the period 2008-2018 is almost non-existent. Even before the FTA, the divergence among importers was falling, and the regression for 2012 still shows the existence of divergence. Therefore, we can conclude that the FTA did not influence substantially the fall in the divergence of importers with respect to share of American imports and the probability of importing goods from the USA.

Table 3.6 Comparison between "frequent" and "lumpy" importers with respect to the share of imports from the USA

	2009	2010	2011	2012 (Year of the FTA)	2013	2014	2015	2016	2017
1[Year> t]* 1[Frequent importer]	0.21***	0.10***	0.06***	0.03***	0.03***	0.02***	0.01***	0.01***	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Observations	321,343	432,498	545,776	664,675	798,556	912,296	1,017,566	1,109,592	1,201,805
R-squared	0.12	0.08	0.05	0.0266	0.03	0.02	0.01	0.01	0.01
Number of firms	29,213	39,318	49,616	60,425	72,596	82,936	92,506	100,872	109,255

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3.7 Comparison between "frequent" and "lumpy" importers with respect to the probability of importing goods from the USA

	2009	2010	2011	2012 (Year of the FTA)	2013	2014	2015	2016	2017
1[Year> t]* 1[Frequent importer]	0.31***	0.17***	0.11***	0.06***	0.04***	0.02***	0.01***	0.00	-0.00**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Observations	321,343	432,498	545,776	664,675	798,556	912,296	1,017,566	1,109,592	1,201,805
R-squared	0.11	0.07	0.05	0.04	0.02	0.02	0.01	0.01	0.01
Number of firms	29,213	39,318	49,616	60,425	72,596	82,936	92,506	100,872	109,255

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Number of country suppliers. Similar to the previous variables, the divergence among existing importers shows a diminishing trend (Table 3.8). This also occurs when we adjust the number of country suppliers by considering accumulation of inventories (Table 3.9). Different than the variables related to imports from the United States, the divergence falls but it does not disappear completely. Moreover, the FTA does not seem to influence significantly the reduction in the divergence among importers, given the fact that this reduction started before the FTA was put in place.

Table 3.8 Comparison between "frequent" and "lumpy" importers with respect to the number of country suppliers

	2009	2010	2011	2012 (Year of the FTA)	2013	2014	2015	2016	2017
1[Year> t]* 1[Frequent importer]	2.06***	1.52***	1.21***	1.01***	0.81***	0.68***	0.53***	0.45***	0.45***
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	321,343	432,498	545,776	664,675	798,556	912,296	1,017,566	1,109,592	1,201,805
R-squared	0.06	0.04	0.03	0.03	0.02	0.01	0.01	0.01	0.00
Number of firms	29,213	39,318	49,616	60,425	72,596	82,936	92,506	100,872	109,255

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3.9 Comparison between "frequent" and "lumpy" importers with respect to the number of country suppliers (adjusted for inventories using the Lu et al (2017) methodology)

	2009	2010	2011	2012 (Year of the FTA)	2013	2014	2015	2016	2017
1[Year> t]* 1[Frequent importer]	2.06***	1.51***	1.21***	1.01***	0.81***	0.68***	0.53***	0.44***	0.45***
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	320,507	431,706	544,962	663,597	797,313	910,877	1,015,960	1,107,920	1,200,034
R-squared	0.06	0.04	0.03	0.03	0.02	0.01	0.01	0.01	0.00
Number of firms	29,137	39,246	49,542	60,327	72,483	82,807	92,360	100,720	109,094

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Number of importing sectors. The divergence between frequent and infrequent importers also falls when we evaluate the variety of the basket of exports, but it does not disappear as it was the case of the regressions related to American imports (Table 3.10). This downward trend in the divergence of importers occurs even before the FTA is put in place. Hence, the trade liberalization does not seem to influence the reduction in the divergence in a substantial way.

Table 3.10 Comparison between "frequent" and "lumpy" importers with respect to the number of sectors the firms purchase foreign goods from

	2009	2010	2011	2012 (Year of the FTA)	2013	2014	2015	2016	2017
1[Year> t]* 1[Frequent importer]	2.52***	1.69***	1.25***	1.01***	0.74***	0.57***	0.39***	0.29***	0.25***
	(0.03)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
Observations	321,343	432,498	545,776	663,597	798,556	912,296	1,017,566	1,109,592	1,201,805
R-squared	0.07	0.05	0.03	0.03	0.02	0.01	0.01	0.01	0.00
Number of firms	29,213	39,318	49,616	60,327	72,596	82,936	92,506	100,872	109,255

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

3.9. Conclusion

While there is work that shows the differences between importers and non-importers, this paper documents patterns that suggest the existence of heterogeneity among importers and churning behavior. Firstly, we observe a continuous entry of new importers and exit of lumpy importers, that is, the existence of churning of importers. Secondly, there exists convergence of new importers with respect to old importers. Thirdly, there is divergence among importers. Specifically, we observe two groups of importers that behave differently: frequent and lumpy importers.

We analyze whether a new free trade agreement between Colombia and the United States impacts these dynamics, and we conclude this is not the case. More specifically, there is no clear evidence that the churning behavior accelerates after the trade liberalization. Although we observe a reduction in the divergence across existing importers, this occurred even before the FTA was put in place. Moreover, the convergence behavior of new importers with respect to old importers does not seem to be altered by the trade liberalization.

Our findings show dynamic behaviors across importers that seem to be resilient to changes in trade policy. These behavioral patterns across importers have implications to understand the productivity gains generated by imports. Moreover, understanding these patterns helps us to understand potential changes in the composition of imports, which could be driven by importers with specific firm characteristics.

Bibliography

[1] Albornoz, Facundo, Héctor F.Calvo Pardo, Gregory Corcos, and Emanuel Ornelas. 2012. "Sequential Exporting." *Journal of International Economics*, 88(1): 17-31.

[2] Alessandria, George, Joseph P. Kaboski, and Virgiliu Midrigan. 2010. "Inventories, Lumpy Trade, and Large Devaluations." *American Economic Review*, 100(5): 2304-39.

[3] Amiti, Mary, and Jozef Konings. 2007. "Trade Liberalization, Intermediate Inputs, and Productivity: Evidence from Indonesia." *American Economic Review*, 97 (5): 1611-1638.

[4] Andrew B. Bernard, Stephen Redding, and Peter K. Schott. 2004. "Comparative Advantage and Heterogeneous Firms." *CEP Discussion Papers dp0643*, Centre for Economic Performance, LSE.

[5] Bernard, Andrew B., J. Bradford Jensen, and Peter K. Schott. 2005. "Importers, Exporters and Multinationals: A Portrait of Firms in the U.S. that Trade Goods." *Producer Dynamics: New Evidence from Micro Data*, University of Chicago Press.

[6] Bernard, Andrew B., J. Bradford Jensen, Stephen J. Redding, and Peter K. Schott. 2009. "The Margins of US Trade." *American Economic Review*, 99(2): 487-93.

[7] Eaton, Jonathan, Samuel Kortum, and Francis Kramarz. 2011. "An Anatomy of International Trade: Evidence From French Firms." *Econometrica*, 79(5): 1453-1498.

[8] Goldberg, Pinelopi K., Amit K. Khandelwal, Nina Pavcnik, Petia Topalova. 2010. "Imported Intermediate Inputs and Domestic Product Growth: Evidence from India." *The Quarterly Journal of Economics*, 125(4): 1727–1767.

[9] Gopinath, Gita and Brent Neiman. 2014. "Trade Adjustment and Productivity in Large Crises." *American Economic Review*, 104(3): 793-831.

[10] Gilbert, Laurien. 2018. "New exporter growth and sequential entry." Mimeo, University of Michigan.

[11] Halpern, Laszlo, Miklos Koren, and Adam Szeidl. 2015. "Imported Inputs and Productivity." *American Economic Review*, 105(12): 3660–3703.

[12] Hsieh, Chang-Tai, and Peter J. Klenow. 2009. "Misallocation and Manufacturing TFP in China and India." *Quarterly Journal of Economics*, 124(4): 1403–48.

[13] Lu, D., A. Mariscal, and L.F. Mejia. 2017. How Firms Accumulate Inputs. Mimeo, University of Rochester.

[14] Morales, Eduardo, Gloria Sheum, and Andrés Zahler. 2019. "Extended Gravity." *Review of Economic Studies*, 86(6): 2668–2712.

[15] Marc J. Melitz. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica*, 71(6):1695-1725.

APPENDIX A

Appendix for Chapter 1

A. Data

The following list contains detailed notes about data. This includes the geospatial dataset as well as the data regarding the calibration of all parameters. Unless otherwise indicated, I use data for 2013 in all cases.

Departments merged or dropped for the analysis

I merge or drop six departments when I take the model to the data

- San Andres y Providencia. The department is an island.
- Leticia. This department trades with the rest of the world exclusively because there is a regional dynamic between two border towns.
- Bogota (merged). The data from Bogota D.C. was merged with Cundinamarca. This gives us a total of 30 departments for the trade model.
- Vaupes, Vichada and Guainia. The states are not connected to the primary road system. Additionally, their international trade flows are small, and these flows are linked to the regional economic activity of the small border towns in Venezuela or Brazil.

Speed values for public-private roads

I assume higher speed values for public-private roads given that the characteristics of the public-private infrastructure projects suggest higher quality for these roads, relative to the standard ones. Such characteristics are publicly available via documents published by the National Agency of Infrastructure, the government office in charge of public-private infrastructure projects. Such documents include the legal contracts with information about design specifications and fines in case of violations by the construction company, as well as inspection documents.

There exists evidence that the Colombian government enforces these contracts, particularly for very expensive projects. Specifically, Alvear-Sanin (2008) documents a legal case in which the Colombian government sued an conglomerate of construction companies for breach of contract (the legal case of Commsa). The Colombian government attempted to impose the largest fine and persisted through different judiciary instances for nine years until a settlement was reached. Hence, it is safe to assume that the quality of public-private roads is higher compared to the standard roads that are directly administered by the Colombian government.

Trade flows

- **Oil exports**. The customs data does not record the department of origin for 55% of mining exports. This data corresponds to shipments with HS2012 codes 2709, 2710 and 2711 (petroleum and oil products). I use production data at a department level from the Colombian public oil company Ecopetrol to define the source of such flows. I assign the export flows without information about the department of origin proportionally to every department that produces oil, according to the production shares.

- **Trade between departments**.

  Agriculture and manufacturing. I use data of the estimated weight for the annual cargo flows from the Transportation Survey of Origin/Destination 2013 from the Ministry of Transportation in Colombia to create a matrix of domestic trade flows. I assume the domestic trade flows are the same for both sectors.

Mining. I use data regarding oil production from the Ministry of Energy and Mines. I assume that only crude oil is domestically traded given that production of coal and oil represent 88% of the output of the mining sector according to the input-output matrix of Colombia created by DANE, for the year 2010. Additionally, coal is mostly exported by Colombia, according to data from the U.S. Energy Information Administration (2019). Therefore, I assume that most of the trade that occurs between departments will be crude oil from the oil fields to the states with refineries.

- **Purchases of location i to itself**.
  - Purchases of the RoW to itself, $\mu_{RoW,RoW}$. I estimated this value using data from WIOD 2013 to obtain $C_{world,final,k}$ and $C_{world,intermediate,k}$ and the customs data of Colombia to obtain this parameter.
  - Purchases of Colombia to itself, μ_{ColCol}. I estimated this using the input-output matrix produced by DANE for the year 2010.
  - Purchases of a department to itself or $\mu_{dd,k}$. I assume this number for the agricultural and manufacturing sectors. For the case of the mining sector, I obtained a proxy of this parameter for every department. To do so, I assume that all the domestic trade of mining is exclusively crude oil from the oil fields to the refineries, given that 88% of the mining production is coal and crude oil according to DANE, and that Colombia does use very little coal for energy consumption (less than 9%) according to the U.S. Energy International Agency (2019).

Trade deficits

- *Trade deficits beween departments and RoW.* I use customs administrative data from DANE for the year 2013.
- *Trade deficits between departments.*

Agriculture and manufacturing. Use data from the Transportation Survey of Origin/Destination 2013 produced by Ministry of Transportation in Colombia. I assume the trade deficits between departments are very small for agriculture and manufacturing, compared to the deficits of departments with the Rest of the world.

Mining. Similar to the way I obtained the trade flows shares, I calculate this variable assuming that domestic trade between departments is mostly crude oil from departments with oil fields to departments with refineries.

Input-output parameters

- Share of value added. Given that global input-output table of WIOD does not have data for Colombia, to estimate the parameter I consider the data for the entire world. This seems feasible given that Colombia is a very small economy, therefore it is likely that the value of this parameter for the world is the same with/without including the Colombian economy.

- Share of sector k in final demand $\beta_{i,k}$.

  · *Rest of the world*. Use final consumption column of the WIOT 2016. Due to constraints in WIOD data, I estimate the parameter for the entire world.

  · *Colombia*. I use input-output table produced by DANE for the year 2010.

Data sources

The following list provides the sources for every variable used in this paper.

1. **WIOD data**. It contains data for all European countries and other major economies. Colombian data is contained in the rest of the world, thus it is not reported individually. See Timmer et al. (2015). I use the input-output table corresponding to the year 2013 (version 2016).

2. **Colombian statistical agency DANE**

   (a) Input-output matrix

   (b) Value added data

   (c) Sectoral GDP data

3. **Colombian statistical agency (DANE)**. Provides the customs administrative data used to estimate trade flows between departments and rest of the world.

4. **Ministry of Transportation of Colombia (Ministerio de Transporte)**.

   (a) Physical maps regarding the primary road system. This allows me to obtain the road distance between Colombian departments. To create the map of 2013, I use as baseline the digital road map created by the National Institute of Roads (INVIAS) for the year 2014.

   (b) Data regarding the estimated weight of the cargo transported between the capitals of Colombian departments.

5. **International Monetary Fund**. Daily data for the exchange rate Colombian peso per dollar.

6. **Ministry of Mines and Energy**. Data on oil production for the year 2013 and the capacity of all refineries in Colombia.

Geospatial data

I obtain information regarding the location of city-ports and capitals of departments via two sources: the main topographic world map generated by ArcGIS software, and coordinates obtained through Google Maps. For some cases, the location of the city-port was assigned to specific coordinates to make sure that the trade costs from a location to itself was normalized to 1. I describe these cases below.

1. All the goods eported via the international bridge of San Miguel are assigned to Puerto Asis in the customs data. For the purpose of the estimation of distances, I use the actual location of the port of San Miguel.

2. I merged the Port of Coveñas and the Port of Cartagena given that they are located in the same city (Cartagena).

3. When the port is located within the city limits, then I situated the capital in the same location as in the port. The cases where this occurs are: Cartagena, Santa Marta, Pereira, Barranquilla and Bogota, .

   The cases where the port of trade is located outside the city limits are: Medellin, Arauca, Cali, Armenia and Bucaramanga.

4. I considered all the goods that are exported via the Port of Coveñas as exported via the Port of Cartagena given that they are located in the same city (Cartagena).

5. I did not use customs data from the ports of Inirida, Leticia, San Andres, Puerto Carreño. This is because the international trade flows of these towns are mainly influenced by the local border regions. For example, the trade flows observed in the port of Leticia, Colombia are mainly driven by regional dynamics between Leticia and Tabaratinga, Brazil.

B. Historical maps

Figure A.1 Map of Colombia's road network in 1938 from the Atlas de Colombia (IGAC, 2002)

Figure A.2 Map of the colonial routes of the Viceroyalty of New Granada available in the Atlas de Colombia(IGAC, 2002)

C. Derivations

Obtaining the expression for trade flows

By solving the firm's problem I obtain the demand of the composite good

$$q^c_{jn,k} = \frac{p^{-\sigma_k}_{jn,k}}{P^{1-\sigma_k}_{n,k}} Q_{n,k}$$

where $P_{n,k}$ is the price of the composite intermediate good and $p_{n,k}$ is the price of the intermediate good in location n.

Given the existence of perfectly competitive markets, the price charged by a firm located in j that sells good of sector k to composite goods firms in location n is

$$p_{jn,k} = \frac{\tau_{jn} c_{j,k}}{A_{j,k}}$$

Plugging this into the equation for the price of the composite intermediate, $P_{j,k}$, I obtain

$$P_{n,k} = \left[\sum_j p^{1-\sigma_k}_{jn,k}\right]^{\frac{1}{1-\sigma_k}} = \left[\sum_j \left(\frac{\tau_{jn} c_{j,k}}{A_{j,k}}\right)^{1-\sigma_k}\right]^{\frac{1}{1-\sigma_k}}$$

To obtain the expression for trade flows, combine the demand of composite good with the price, to get

$$x_{jn,k} = p_{jn,k} \cdot q^c_{jn,k} = \frac{p^{1-\sigma_k}_{jn,k}}{P^{-\sigma_k}_{nk}} Q_{n,k} \iff$$

$$X_{jn,k} = \left(\frac{\tau_{jn,k} c_{j,k}}{A_{j,k}}\right)^{1-\sigma_k} Q_{n,k} P^{\sigma_k - 1}_{n,k}$$

Obtaining the labor market clearing

By aggregating the total expenditure of location n in sector-k goods (14) across all sectors, I obtain the total expenditure of location n

$$X_n = \sum_s X_{n,s} = \sum_s \Big[\sum_k \Big(\beta_n^{s,k} \sum_j X_{j,k} \lambda_{nj,k} \Big) + \alpha_{n,s} I_n \Big]$$

$$M_n = X_n = \sum_s \sum_k \beta_n^{s,k} \sum_j X_{j,k} \lambda_{nj,k} + w_n L_n + D_n$$

$$E_n = \sum_k \sum_j X_{j,k} \lambda_{nj,k} = M_n - D_n = \sum_k (1 - \beta_n^{l,k}) \sum_j X_{j,k} \lambda_{nj,k} + w_n L_n$$

where he first equality comes from the trade balance equation. After some algebra, I can obtain an expression for labor market clearing.

$$w_n L_n = \sum_k \beta_n^{l,k} \sum_j X_{nj,k} = \sum_k \beta_n^{l,k} \sum_j X_{j,k} \lambda_{nj,k}$$

Definition of equilibrium in levels (detailed).

The equilibrium is a set of wages $\{w_{n,k}\}_{n \in Z, k \in \{a,m,i\}}$*, prices* $\{P_{n,k}\}_{n \in Z, k \in \{a,m,i\}}$*, and labor allocations* $\{L_{n,k}\}_{n \in Z, k \in \{a,m,i\}}$ *for all locations* $n \in Z$ *under the assumption of labor mobility across sectors and immobile labor across locations, given the following parameters:*

(a) *trade costs* $\{\tau_{ij}\}_{n,j \in R}$,

(b) *share of value added of sector* s *in the production of sector* k $\{\beta_n^{s,k}\}_{n \in R, s,k \in \{a,m,i,z\}}$,

(c) *elasticity of substitution* $\{\sigma_k\}_{k \in \{a,m,i,z\}}$,

(d) *labor endowments* $\{L_n\}_{n \in R}$,

(e) *and total trade deficits* $\{D_n\}_{n \in R}$

that solve the following system of equations:

i *Wages.*

$$w_i = w_{i,k} \forall k$$

ii *Cost of an input bundle*

$$c_{n,k} = \phi_{n,k}(w_n)^{\beta_n^{l,k}} \prod_{s\in\{a,m,i,z\}} (P_{n,s})^{\beta_n^{s,k}}$$

iii *Prices.*

$$P_{n,k} = \Big[\sum_j \Big(\frac{\tau_{jn}c_{j,k}}{A_{j,k}}\Big)^{1-\sigma_k}\Big]^{\frac{1}{1-\sigma_k}}$$

iv *Trade flows shares.*

$$\lambda_{jn,k} = (\tau_{jn})^{1-\sigma_k}(c_{j,k})^{1-\sigma_k}(P_{n,k})^{\sigma_k-1}{A_{j,k}}^{\sigma_k-1}$$

v *Total expenditure.*

$$X_{n,s} = \sum_k \beta_n^{s,k} \sum_j X_{j,k}\lambda_{nj,k} + \alpha_{n,s}I_n$$

where

$$I_n = w_nL_n + D_n$$

vi *Trade balance*[18].

$$\sum_k \sum_{j\in R} X_{j,k}\lambda_{nj,k} = \sum_k \sum_{j\in R} X_{n,k}\lambda_{jn,k} - D_n$$

Transportation framework

Probability that the shipping cost offer is lower than c

Consider a shipping route $r_t \in R_t$ for $t \in \{x,m\}$. Denote the potential shipping cost of a trader ι as $\tau^o_{r_t,k}$. This offer depends on the shipping cost along route r_t and a productivity draw $z_{r_t,k}(\iota)$, which follows a Frechet distribution with parameters $(A_{r_t,k}\theta_k)$.

[18]This condition implies labor market clearing
$w_nL_n = \sum_k \beta_n^{l,k} \sum_{j\in R} X_{j,k}\lambda_{nj,k}$

$$\tau^{o}_{r_t,k}(\iota) = \frac{\tau_{r_t,k}}{z_{r_t,k}(\iota)}$$

It can be noticed that the higher the value of the draw, the lower the shipping cost offer along route r_t. The probability that the shipping cost offer is lower than c is given by

$$G_{r_t,k}(c) = Pr\Big[\tau^{o}_{r_t,k}(\iota) \leq c\Big] \iff$$

$$G_{r_t,k}(c) = Pr\Big[\frac{\tau_{r_t,k}}{z_{r_t}(\iota)} \leq c\Big] \iff$$

$$G_{r_t,k}(c) = Pr\Big[z_{r_t}(\iota) \geq \frac{\tau_{r_t,k}}{c}\Big] \iff$$

$$G_{r_t,k}(c) = 1 - Pr\Big[z_{r_t}(\iota) \leq \frac{\tau_{r_t,k}}{c}\Big] \iff$$

$$G_{r_t,k}(c) = 1 - F\Big(\frac{\tau_{r_t,k}}{c}\Big) \iff$$

$$G_{r_t,k}(c) = 1 - exp[-A_{r_t}(\tau_{r_t})^{-\theta_k} c^{\theta_k}]$$

Let $\tau_s(\iota)$ be the actual shipping cost of trader ι from department d to the rest of the world. This cost is the minimum shipping price among all potential shipping cost offers across city-ports, that is

$$\tau_s(\iota) = \underset{r_t}{min}\ \tau^{o}_{r_t,k}(\iota) = \underset{r_t}{min}\ \frac{\tau_{r_t,k}}{z_{r_t,k}(\iota)}$$

Probability that the observed shipping cost is lower than c

Let $G_{t,k}(c)$ be the probability that the *observed* shipping cost $\tau_s(\iota)$ is lower than c. Therefore, I have

$$G_{t,k}(c) \equiv Pr\Big[\tau_s(\iota) \leq c\Big] = Pr\Big[\underset{r_t}{min}\frac{\tau_{r_t,k}}{z_{r_t,k}(\iota)} \leq c\Big] \iff$$

$$G_{t,k}(c) = 1 - Pr\Big[\underset{r_t}{min}\frac{\tau_{r_t,k}}{z_{r_t,k}(\iota)} \geq c\Big]] \iff$$

$$G_{t,k}(c) = 1 - Pr\Big[\underset{r_t \in R_t}{\cap}\frac{\tau_{r_t,k}}{z_{r_t,k}(\iota)} \geq c\Big] \iff$$

$$G_{t,k}(c) = 1 - Pr\Big[\underset{r_t \in R_t}{\cap}\frac{\tau_{r_t,k}}{z_{r_t,k}(\iota)} \geq c\Big] \iff$$

$$G_{t,k}(c) = 1 - \prod_{r_t \in R_t} Pr\Big[\frac{\tau_{r_t,k}}{z_{r_t,k}(\iota)} \geq c\Big] \iff$$

$$G_{t,k}(c) = 1 - \prod_{r_t \in R_t} [1 - G_{r_t}(c)]$$

Plugging the expression $G_{r_t}(c) = 1 - exp[-A_{r_t}(\tau_{r_t})^{-\theta_k}c^{\theta_k}]$ into the previous equation, I obtain

$$G_{t,k}(c) = 1 - \prod_{r_t \in R_t} exp\Big[-A_{r_t}(\tau_{r_t})^{-\theta_k}c^{\theta_k}\Big] \iff$$

$$G_{t,k}(c) = 1 - exp\Big[-c^{\theta_k}\sum_{r_t} A_{r_t}(\tau_{r_t})^{-\theta_k}\Big] \iff$$

$$G_{t,k}(c) = 1 - exp\Big[-c^{\theta_k}\Phi_t\Big]$$

where $\Phi_t \equiv \sum_{r_t} A_{r_t}(\tau_{r_t})^{-\theta_k}$.

Probability that any good is shipped via route r_t

Denote $\pi_{r_t,k}$ the probability that any good is shipped via route $r_t \in R_t$. Similar to Eaton and Kortum (2002), given that specialized traders have i.i.d. draws that are sector k specific in my framework, then $\pi_{r_t,k}$ is also the fraction of goods of sector k that are shipped via route r_t.

$$\pi_{r_t,k} \equiv Pr\Big[\tau^o_{r_t,k}(\iota) \leq \underset{v_t \in R_t \setminus r_t}{min} \tau^o_{v_t,k}(\iota)\Big] \iff$$

$$\pi_{r_t,k} = \int_0^\infty Pr\Big[\underset{v_t \in R_t \setminus r_t}{min} \tau^o_{v_t,k}(\iota) \geq c\Big] dG_{r_t,k}(c) \iff$$

$$\pi_{r_t,k} = \int_0^\infty Pr\Big[\bigcap_{v_t \in R_t \setminus r_t} \{\tau^o_{v_t,k}(\iota) \geq c\}\Big] dG_{r_t,k}(c) \iff$$

$$\pi_{r_t,k} = \int_0^\infty \prod_{v_t \in R_t \setminus r_t} [1 - G_{v_t}(c)] dG_{r_t,k}(c) \iff$$

$$\pi_{r_t,k} = \int_0^\infty \prod_{v_t \in R_t \setminus r_t} [1 - G_{v_t}(c)] dG_{r_t,k}(c) \iff$$

Using the expressions $G_{v_t,k}(c) = 1 - exp[-A_{v_t}(\tau_{v_t})^{-\theta_k} c^{\theta_k}]$, and $dG_{r_t,k}(c) = \frac{d}{dc}[1 - exp(-A_{r_t}(\tau_{r_t})^{-\theta_k} c^{\theta_k})]dc$, I obtain

$$\pi_{r_t,k} = \int_0^\infty \prod_{v_t \in R_t \setminus r_t} \Big[exp\Big(-A_{v_t}(\tau_{v_t})^{-\theta_k} c^{\theta_k}\Big)\Big] \Big[\frac{d}{dc}[1 - exp(-A_{r_t}(\tau_{v_t})^{-\theta_k} c^{\theta_k})]\Big] dc \iff$$

$$\pi_{r_t,k} = A_{r_t}(\tau_{r_t})^{-\theta_k} \int_0^\infty \theta_k c^{\theta_k - 1}[exp(-c^{\theta_k}\Phi_t)]dc$$

$$\pi_{r_t,k} = \frac{A_{r_t}(\tau_{r_t})^{-\theta_k}}{\Phi_t}\Big[-exp(-c^{\theta_k}\Phi_t|_0^\infty)\Big]$$

$$\pi_{r_t,k} = \frac{A_{r_t}(\tau_{r_t})^{-\theta_k}}{\Phi_t}$$

Why π_{r_t} is the fraction of trade flows between department d and the rest of the world that are shipped via route r_t

So far, I have shown that π_{r_t} is the fraction of exports/imports by department d to/from the rest of the world, RoW. But this is not the same as the percentage of the value of trade flows shipped via route r_t. Hence, I need to show that the distribution of shipping cost offers is independent of the shipping route. If this is true, then I can consider r_t as the fraction of exports/imports shipped via route r_t.

I express the probability that the shipping cost offer is lower than $\bar{c}$ conditional on route r_t offering the lowest price as

$$Pr[\tau^o_{r_t}(\iota) \leq \bar{c} | \tau^o_{r_t}(\iota) \leq \min_{v_t \in R_t \backslash r_t} \tau^o_{v_t}(\iota)] = \frac{1}{\pi_{r_t}} \int_0^{\bar{c}} Pr[\min_{v_t \in R_t \backslash r_t} \tau^o_{v_t}(\iota) \geq c] dG_{r_t}(c)$$

$$= \frac{1}{\pi_{r_t}} \int_0^{\bar{c}} \prod_{v_t \in R_t \backslash r_t} [1 - G_{v_t}(c)] dG_{r_t}(c)$$

Combining $G_{v_t,k}(c)$ and $dG_{r_t,k}(c)$ with my last expression, I get

$$Pr\Big[\tau^o_{r_t}(\iota) \leq \bar{c} | \tau^o_{r_t}(\iota) \leq \min_{v_t \in R_t \backslash r_t} \tau^o_{v_t}(\iota)\Big] = \int_0^{\bar{c}} \prod_{v_t \in R_t \backslash r_t} [exp(-A_{v_t}\tau_{v_t})] \frac{d}{dc}[1 - exp(-A_{r_t}(\tau_{r_t})^{-\theta_k} c^{\theta_k})] dc$$

$$Pr\Big[\tau^o_{r_t}(\iota) \leq \bar{c} | \tau^o_{r_t}(\iota) \leq \min_{v_t \in R_t \backslash r_t} \tau^o_{v_t}(\iota)\Big] = \frac{1}{\pi_{r_t}} \frac{A_{r_t}(\tau_{r_t})^{-\theta_k}}{\Phi_t} \Big[- exp(-c^{\theta_k}\Phi_t|^{\bar{c}}_0) \Big]$$

$$Pr\Big[\tau^o_{r_t}(\iota) \leq \bar{c} | \tau^o_{r_t}(\iota) \leq \min_{v_t \in R_t \backslash r_t} \tau^o_{v_t}(\iota)\Big] = \frac{1}{\pi_{r_t}} \frac{A_{r_t}(\tau_{r_t})^{-\theta_k}}{\Phi_t} \Big[1 - exp(-\bar{c}^{\theta_k}\Phi_t) \Big]$$

$$Pr\Big[\tau^o_{r_t}(\iota) \leq \bar{c} | \tau^o_{r_t}(\iota) \leq \min_{v_t \in R_t \backslash r_t} \tau^o_{v_t}(\iota)\Big] = G_{t,k}(\bar{c})$$

The distribution of shipping cost offers is the same for department d, independently of the route r_t used to transport the good. Therefore, the average value of the shipment sold/purchased by department d is independent of the route taken. This implies that we can express the fraction of the value of exports/imports that use shipping route r_t as π_{r_r}. This intuition is similar to the intuition of the result of Eaton and Kortum (2002), the best routes are more efficient, therefore such routes transport a larger share of goods to/from department d from/to the rest of the world, up to the level where the shipping cost offers are equal to the distribution of the observed shipping costs.

Trade costs between a department and the rest of the world

Using the results of the model with traders of Allen and Arkolakis (2019), define the trade cost between a department d and the rest of the world, as

$$\tau_{dRoW} \equiv E\Big[\tau_s(\iota)\Big]$$

$$\tau_{dRoW} = \int_0^\infty p_s(\iota) \iff$$

$$\tau_{dRoW} = \int_0^\infty p\ dG_t(p) \iff$$

$$\tau_{dRoW} = \int_0^\infty pdG_t(p) \iff$$

$$\tau_{dRoW} = \int_0^\infty p\ \frac{d}{dp}[1 - exp(-p^\theta \Phi_t)]dp \iff$$

$$\tau_{dRoW} = \int_0^\infty p\frac{d}{dp}[1 - exp(-p^\theta \Phi_t)]dp \iff$$

$$\tau_{dRoW} = \int_0^\infty \theta \Phi_t p p^{\theta-1} exp(-p^\theta \Phi_t)dp$$

Now, use change of variables, where $x = p^\theta \Phi_t$ and $dx = \theta p^{\theta-1}\Phi_t$. Therefore, I can express the integral as

$$\tau_{dRoW} = \int_0^\infty \left(\frac{x}{\Phi_t}\right)^{\frac{1}{\theta}} exp(-x)dx \iff$$

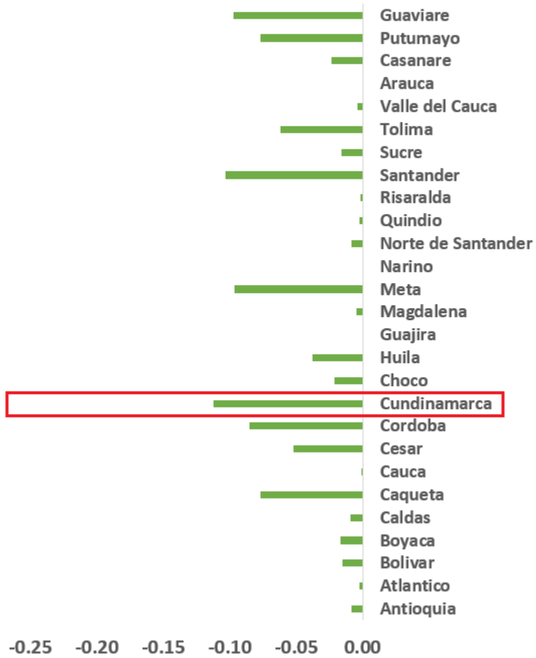
$$\tau_{dRoW} = \Phi_t^{-\frac{1}{\theta}} \int_0^\infty x^{\frac{1}{\theta}} e^{-x} dx$$

Recall that $\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx$. If I consider $(t-1) = \frac{1}{\theta} \iff t = \frac{1+\theta}{\theta}$, then I can express the trade cost between a department d and the rest of the world as

$$\tau_{dRoW} = \Phi_t^{\frac{1}{\theta}} \Gamma\Big(\frac{1+\theta}{\theta}\Big)$$

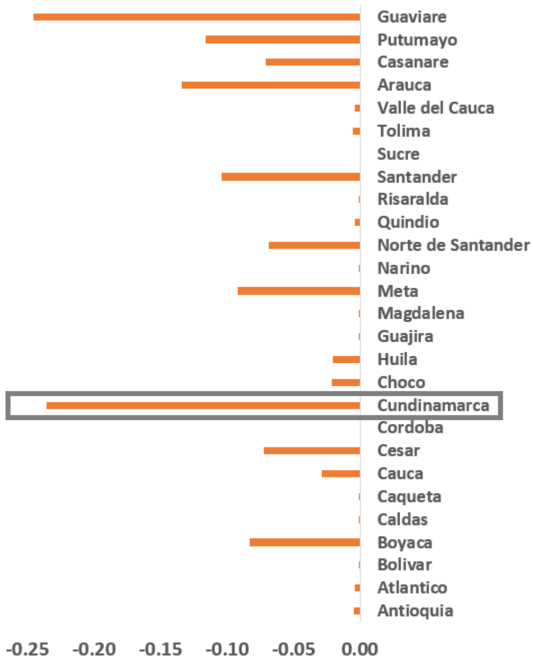
D. Reductions in trade costs $\tau_{dRoW,k}$ generated by the completion of Ruta del Sol

Figure A.3 Reductions of $\tau_{dRoW,agriculture}$ caused by Ruta del Sol



Note: I simulate the change in international trade costs using the value of β_{time} from Allen and Arkolakis (2019).

Figure A.4 Reductions of $\tau_{dRoW,agriculture}$ caused by Ruta del Sol



Note: I simulate the change in international trade costs using the value of β_{time} from Allen and Arkolakis (2019).

Figure A.5 Reductions of $\tau_{dRoW,agriculture}$ caused by Ruta del Sol

Note: I simulate the change in international trade costs using the value of β_{time} from Allen and Arkolakis (2019).

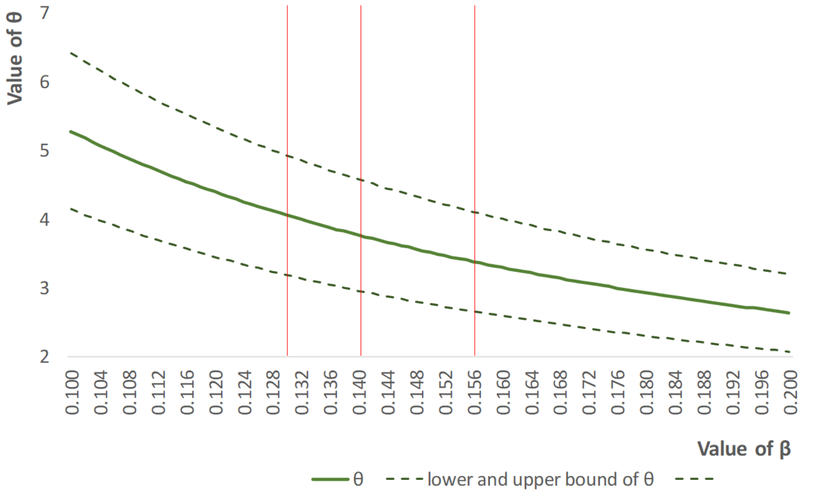
E. Robustness checks for the simulations of Ruta del Sol

Table A.1 Effects of the "Ruta del Sol" infrastructure project in sectoral exports of Colombia under different values of β_{time}

Scenario	$\bar{\mu}_{dd}$	Value of β_t that defines impact of project on τ_{dRoW}	$X_{agr.}/X_{total}$	X_{mining}/X_{total}	$X_{manuf.}/X_{total}$	**Δ share of manufacturing exports**
	0.3	No project (baseline scenario)	7.70 %	54.19%	38.10%	
A		0.13	7.53 %	50.15 %	42.32 %	**+4.22**
B		0.143	7.14 %	49.13 %	43.73 %	**+5.63**
C		0.156	6.9 %	47.3 %	45.8 %	**+7.72**
	0.6	No project (baseline scenario)	7.39%	55.76 %	36.85 %	
D		0.13	7.43 %	51.37 %	41.21 %	**+4.35**
E		0.143	7.09 %	50.41 %	42.51 %	**+5.65**
F		0.156	7.05 %	48.44 %	44.51 %	**+7.65**

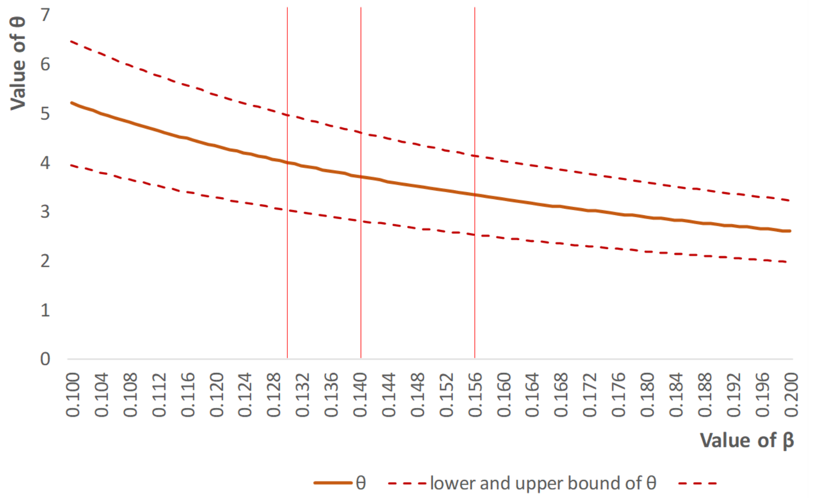
F. Values of θ_k considering the confidence intervals of $\hat{\mu_k}$

Figure A.6 Value of parameter θ_a when considering the confidence interval of $\hat{\mu_a}$



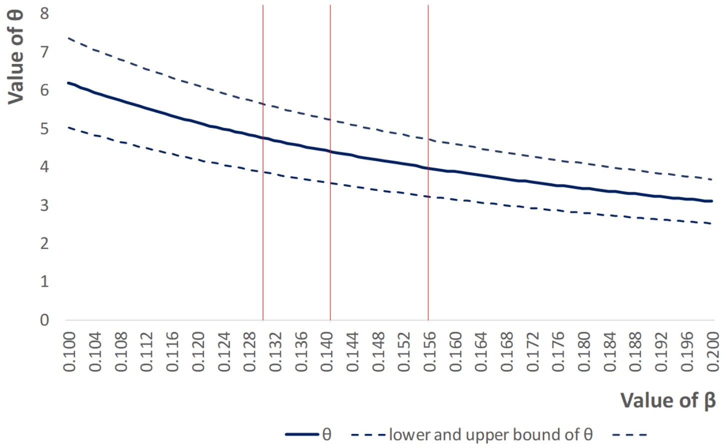
Notes: I use the estimate of $\hat{\mu_a}$ obtained using 2SLS. See Table 2.

Figure A.7 Value of parameter θ_m when considering the confidence interval of $\hat{\mu_m}$



Notes: I use the estimate of $\hat{\mu_m}$ obtained using 2SLS. See Table 2.

Figure A.8 Value of parameter θ_i when considering the confidence interval of $\hat{\mu}_i$



Notes: I use the estimate of $\hat{\mu}_i$ obtained using 2SLS. See Table 2.

G. Simulated change in trade costs after the completion of the highway "Ruta del Sol"

Figure A.9 Simulated change in trade costs before/after Rutal del Sol is finished for agricultural sector

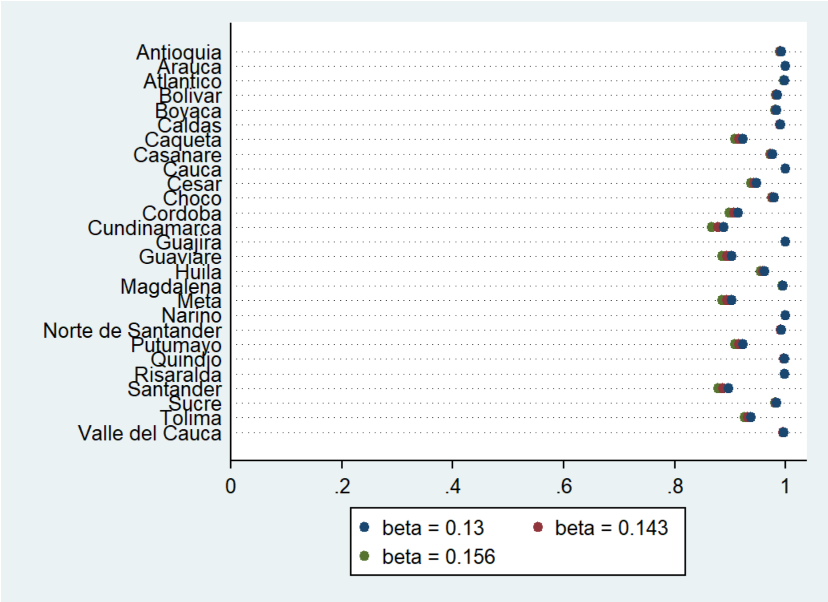


Figure A.10 Simulated change in trade costs before/after Rutal del Sol is finished for mining sector

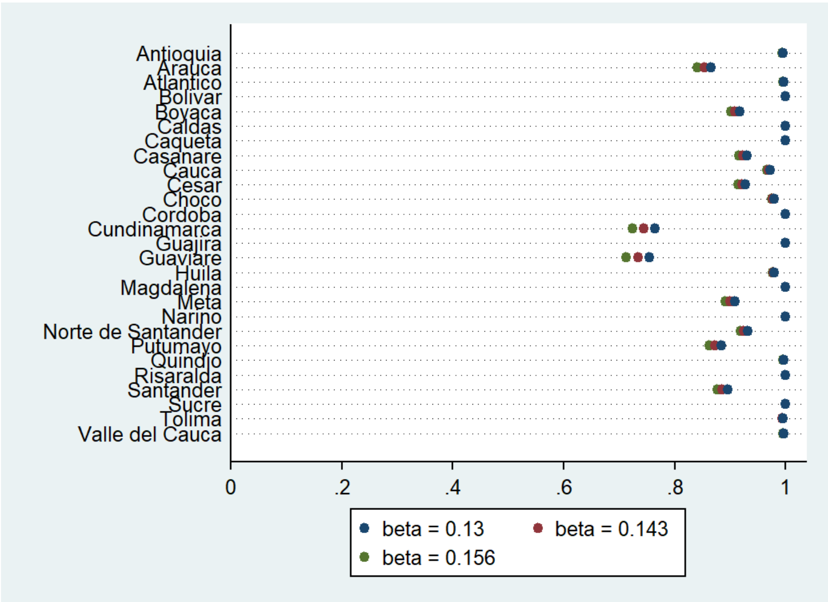
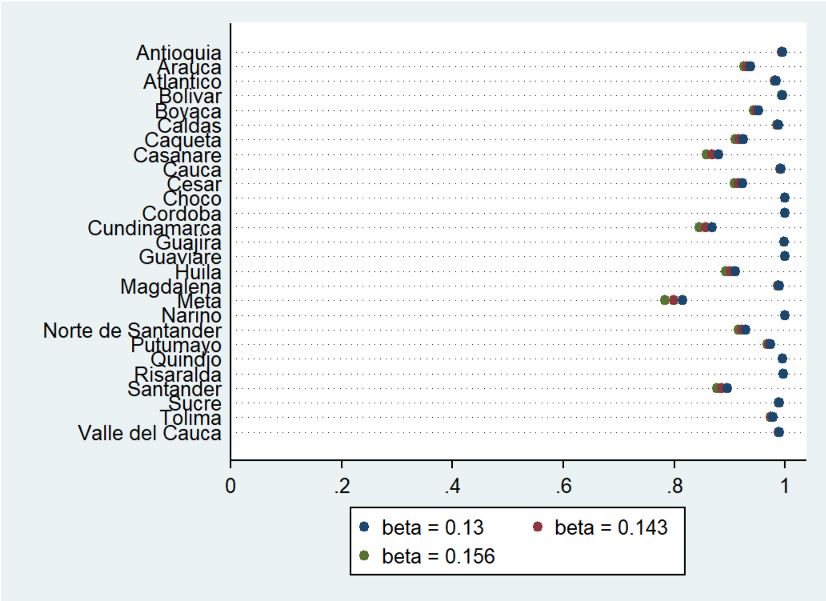


Figure A.11 Simulated change in trade costs before/after Rutal del Sol is finished for manufacturing sector



APPENDIX B

Appendix for Chapter 2

A. Groups of industries and HS 2-digit codes

Table B.1 Groups of industries and ISIC version 3 Adjusted for Colombia (2-digit) code

Group of industries	**2-digit code of ISIC 3 Adjusted for Colombia**	**Description ISIC 3 Adjusted to Colombia**
Food, Beverage and Tobbaco	15	Food products and beverages
	16	Tobbaco products
Wood, paper, textiles, leather	17	Textiles
	18	Clothing, tanning and dying of animal skin.
	19	Tanning and dying leather; shoes, travel accessories, suitcases, bags, and similar goods; saddlery.
	20	Wood manufacturing (including furniture), basketry, plaiting
	21	Paper and cardboard products
	22	Printing, edition, publicity
Chemical industries	23	Coking, oil refining, and nuclear fuel
	24	Chemical substances and chemical products
	25	Rubber and plastics products
	26	Non-metallic mineral products (glass, cercamics, cement, etc.)
	27	Basic metallurgy
Electrical, mechanical and transportation industries	28	Metal products except machinery and equipment
	29	Metal machinery and equipment
	30	Office, accounting and informatic equipment
	31	Electrical machinery and equipment
	32	Manufacturing of radio, tv and telecommunications devices and equipment
	33	Installing, maintenance and repairment of machinery and equipment
	34	Cars and trucks
	35	Manufacturing of other types of transportation

Table B.2 Groups of industries and ISIC version 4 Adjusted for Colombia (2-digit) code

Group of industries	**2-digit code of ISIC 4 Adjusted for Colombia**	**ISIC 4 Adjusted for Colombia**
Food, Beverage and Tobbaco	10	Food products
	11	Beverages
	12	Tobbaco products
Wood, paper, textiles, leather	13	Textiles
	14	Clothing
	15	Animal skin processing; leather processing; manufacturing of travel accessories, bags, luggage, and similar products; saddlery.
	16	Wood manufacturing (except furniture); basketry, plaiting.
	17	Paper and cardboard manufacturing
	18	Printing activities
	31	Furniture
	58	Edition activities
	73	Publicity
Chemical industries	19	Coking, oil refining, fuel production
	20	Chemical substances and chemical products
	21	Pharmaceuticals, chemical products for medicine, botanical products for pharmaceutical use
	22	Rubber and plastics products
	23	Non-metallic mineral products (glass, cercamics, cement, etc.)
	24	Basic metallurgy
Electrical, mechanical and transportation industries	25	Metal products except machinery and equipment
	26	Informatic, electronic and optical equipment
	27	Electrical devices and equipment
	28	Metal machinery and equipment
	29	Cars and trucks
	30	Manufacturing of other types of transportation
	33	Installment, maintenance and reparation of machinery and equipment

B. Estimates of impacts of oil production in different outcomes (real values)

Table B.3 Estimates of the elasticities of industrial production (real values) with respect to oil production, by sector

Dependent variable: industrial production (real values)					
	Food beverage tobaccco	Wood, paper textiles leather	Chemical sector	Electrical machinery transportation	Other industries
log(oil production)	0.76***	-1.22***	1.00***	-1.06***	0.08
	(0.26)	(0.23)	(0.20)	(0.19)	(0.08)
Observations	480	480	480	480	480
R-squared	0.69	0.38	0.23	0.17	0.04
Number of departments	32	32	32	32	32

Standard errors in parentheses
*** p¡0.01, ** p¡0.05, * p¡0.1

Table B.4 Estimates of the elasticities of real wages with respect to oil production, by sector

Dependent variable: average wage by department (real values)					
	Food beverage tobaccco	Wood, paper textiles leather	Chemical sector	Electrical machinery transportation	Other industries
log(oil production)	0.44***	-0.62***	0.61***	-0.64***	0.01
	(0.13)	(0.13)	(0.12)	(0.12)	(0.04)
Observations	480	480	480	480	480
R-squared	0.69	0.31	0.18	0.14	0.03
Number of departments	32	32	32	32	32

Standard errors in parentheses
*** p¡0.01, ** p¡0.05, * p¡0.1

Table B.5 Estimates of elasticities of gross investment (real values) with respect to local oil production, by sector

Dependent variable: gross investment (real values)					
	Food beverage tobaccco	Wood, paper textiles leather	Chemical sector	Electrical machinery transportation	Other industries
log(oil production)	0.66***	-1.02***	0.72***	-0.32	-0.22
	(0.23)	(0.27)	(0.19)	(0.23)	(0.26)
Observations	480	477	479	479	479
R-squared	0.66	0.26	0.18	0.08	0.05
Number of departments	32	32	32	32	32

Standard errors in parentheses
*** p¡0.01, ** p¡0.05, * p¡0.1

Table B.6 Estimates of the elasticities of fixed assets (real values) with respect to oil production, by sector

Dependent variable: fixed assets (real values)					
	Food beverage tobaccco	Wood, paper textiles leather	Chemical sector	Electrical machinery transportation	Other industries
log(oil production)	0.45***	-1.19***	0.98***	-1.04***	0.00
	(0.17)	(0.22)	(0.20)	(0.18)	(0.08)
Observations	480	480	480	480	480
R-squared	0.26	0.37	0.22	0.18	0.03
Number of departments	32	32	32	32	32

Standard errors in parentheses
*** p¡0.01, ** p¡0.05, * p¡0.1

Table B.7 Estimates of the elasticities of the value of intermediate inputs (real values) with respect to oil production, by sector

Dependent variable: value of intermediate inputs (real values)					
	Food beverage tobaccco	Wood, paper textiles leather	Chemical sector	Electrical machinery transportation	Other industries
log(oil production)	0.74***	-1.16***	0.96***	-1.00***	0.08
	(0.26)	(0.22)	(0.20)	(0.18)	(0.08)
Observations	480	480	480	480	480
R-squared	0.69	0.39	0.23	0.17	0.04
Number of departments	32	32	32	32	32

Standard errors in parentheses
*** p¡0.01, ** p¡0.05, * p¡0.1

Table B.8 Estimates of elasticities of purchases of foreign inputs (real values) with respect to oil production, by sector

Dependent variable: purchases of foreign inputs (real values)					
	Food beverage tobaccco	Wood, paper textiles leather	Chemical sector	Electrical machinery transportation	Other industries
log(oil production)	0.84***	-1.06***	0.08	-0.18	-0.01
	(0.27)	(0.20)	(0.16)	(0.14)	(0.19)
Observations	480	480	480	480	480
R-squared	0.35	0.13	0.04	0.06	0.09
Number of departments	32	32	32	32	32

Standard errors in parentheses
*** p¡0.01, ** p¡0.05, * p¡0.1

Table B.9 Estimates of elasticities of the expenditures of transportation inputs (real values) with respect to oil production, by sector

Dependent variable: expenditures on transportation of inputs (real values)					
	Food beverage tobaccco	Wood, paper textiles leather	Chemical sector	Electrical machinery transportation	Other industries
log(oil production)	0.51***	-1.00***	0.75***	-0.69***	0.10
	(0.18)	(0.17)	(0.15)	(0.14)	(0.09)
Observations	480	480	480	480	480
R-squared	0.39	0.44	0.20	0.24	0.21
Number of departments	32	32	32	32	32

Standard errors in parentheses
*** p¡0.01, ** p¡0.05, * p¡0.1

APPENDIX C

Appendix for Chapter 3

A. Groups of industries and HS 2-digit codes

Table C.1 Groups of industries and ISIC version 3 Adjusted for Colombia (2-digit) code

Group of industries	**2-digit code of ISIC 3 Adjusted for Colombia**	**Description ISIC 3 Adjusted to Colombia**
Food, Beverage and Tobbaco	15	Food products and beverages
	16	Tobbaco products
Wood, paper, textiles, leather	17	Textiles
	18	Clothing, tanning and dying of animal skin.
	19	Tanning and dying leather; shoes, travel accessories, suitcases, bags, and similar goods; saddlery.
	20	Wood manufacturing (including furniture), basketry, plaiting
	21	Paper and cardboard products
	22	Printing, edition, publicity
Chemical industries	23	Coking, oil refining, and nuclear fuel
	24	Chemical substances and chemical products
	25	Rubber and plastics products
	26	Non-metallic mineral products (glass, cercamics, cement, etc.)
	27	Basic metallurgy
Electrical, mechanical and transportation industries	28	Metal products except machinery and equipment
	29	Metal machinery and equipment
	30	Office, accounting and informatic equipment
	31	Electrical machinery and equipment
	32	Manufacturing of radio, tv and telecommunications devices and equipment
	33	Installing, maintenance and repairment of machinery and equipment
	34	Cars and trucks
	35	Manufacturing of other types of transportation

Table C.2 Groups of industries and ISIC version 4 Adjusted for Colombia (2-digit) code

Group of industries	**2-digit code of ISIC 4 Adjusted for Colombia**	**ISIC 4 Adjusted for Colombia**
Food, Beverage and Tobbaco	10	Food products
	11	Beverages
	12	Tobbaco products
Wood, paper, textiles, leather	13	Textiles
	14	Clothing
	15	Animal skin processing; leather processing; manufacturing of travel accessories, bags, luggage, and similar products; saddlery.
	16	Wood manufacturing (except furniture); basketry, plaiting.
	17	Paper and cardboard manufacturing
	18	Printing activities
	31	Furniture
	58	Edition activities
	73	Publicity
Chemical industries	19	Coking, oil refining, fuel production
	20	Chemical substances and chemical products
	21	Pharmaceuticals, chemical products for medicine, botanical products for pharmaceutical use
	22	Rubber and plastics products
	23	Non-metallic mineral products (glass, cercamics, cement, etc.)
	24	Basic metallurgy
Electrical, mechanical and transportation industries	25	Metal products except machinery and equipment
	26	Informatic, electronic and optical equipment
	27	Electrical devices and equipment
	28	Metal machinery and equipment
	29	Cars and trucks
	30	Manufacturing of other types of transportation
	33	Installment, maintenance and reparation of machinery and equipment